Empowering Audiences - Against Misinformation Through 'Prebunking'

Research-Based Insights on the Problem of Misinformation and Steps Towards Its Solution

Michael Bang Petersen
Professor of Political Science
Aarhus University
Denmark

June 1, 2023
Background report commissioned by *The Future of Free Speech.*

Executive Summary

- Exposure to misinformation can change people's views but often the impact is limited because specific attitudes are often rooted in larger worldviews that are difficult to change. Nonetheless, exposure to misinformation can have a number of other adverse consequences. It can distract people's time and attention from more relevant information sources; it can deplete politician's and mainstream media's resources because of a constant felt need to counter it; and, even if it does not change views, it can still sow confusion and uncertainty.

- A small minority of users are responsible for the spread of most misinformation – the habitual "super-sharers". These users are motivated by political activism rooted in difficult-to-change psychological dispositions such as anger and frustration against specific political groups and actors.

- Simple online interventions that target "super-sharers" re likely to fail. Dealing with these individuals will likely require genuine deradicalization programs targeting current "super-sharers" as well as actual policy reform to address the underlying frustrations motivating new "super-sharers" to enter social media.

- Educational interventions are mainly likely to succeed if they focus on empowering the online audience who occasionally is be exposed to misinformation as consequence of the activities of the "super-sharers". This involves a conceptual shift from focusing interventions on changing the motivations of habitual sharers to mitigating the effects of their actions on naive bystanders.

- One well-known intervention is fact-checking. Fact-checking works at two levels: First, it incentivizes decision-makers to share more accurate information and, second, exposure to fact-checks reduces belief in and the sharing of misinformation. Yet, the effectiveness of fact-checks – sometimes referred to as debunking - is reduced by the fact that it is necessarily reactive rather than proactive (so-called prebunking); that misinformation producers works fast and in unexpected ways; and that research shows that it requires repeated exposure to a fact-check to keep false beliefs from emerging.

- A promising avenue is therefore to invest in prebunking interventions. These interventions are designed to (1) foster the awareness of misinformation and (2)

develop competences in detecting misinformation have been found to be successful and in reducing the sharing of misinformation without necessarily reducing the sharing of valid information. Such prebunking interventions are thus oriented towards empowering audiences. An orientation towards empowerment is in line with general principles in risk communication research about how to motivate the public to deal with risks across domains.

- Prebunking interventions should focus on establishing intellectual humility. Trust and humility are positively related to the sharing of and belief in reliable information. Generalized mistrust, in contrast, is positively related to the sharing of misinformation. It is key that the interventions not only focus on facilitating a critical and suspicious mindset. Instead, interventions need to highlight the fallibility of users' own intuitions and thereby foster humility.

Contents

Introduction

A central focus of public debate is the spread of false information and news on social media. The public is concerned according to polls (Newman etla., 2018). A report from Reuters Institute shows that in 2018, 54 % across 37 countries from all over the world was on average "concerned about what is real and what is fake on the internet". People in Brazil and Portugal are most concerned, with 85 and 71 % being concerned, respectively. People in the Netherlands and Denmark was least concerned, with 30 and 36 % being concerned respectively. During the COVID-19 pandemic, these concerns was intensified globally, as the World Health Organization warned about the associated "infodemic" of information including false information during the crisis situation (Briand et al., 2021).

The concerns over the circulation of false information have led to a range of initiatives. These include political regulation of social media platforms to hold them accountable for the circulation; social media platforms' own attempts to downregulate the spread of misinformation by changing algorithms and flagging false content; and educational initiatives directed towards users to decrease the likelihood that they believe and share false information when they encounter it. These initiatives range from short instructive messages that are can be met online on social media to long educational programs that can be completed in educational settings such as schools.

This background report focuses on set of such initiatives: Short and scalable user-directed educational initiatives that can be fielded online. In particular, the background report assesses the evidence in favor of interventions that seek to empower online audiences to become their own fact-checkers and learn after to discern between true and false information on social media. This is referred to as *prebunking* initiatives as opposed to more traditional *debunking* initiatives (Lewandowsky & Van Der Linden, 2021). With debunking, concrete news story are debunked by professional fact-checkers and these fact-checks are then broadcasted to users via traditional and social media. With prebunking, users learn general principles about how to tell truth from falsehood and, hence, are ready to react when any concrete news story appear in their social media feed.

While prebunking initiatives cannot solve the problem of misinformation on their own, they are cost-effective and avoid the epistemological and normative challenges related to topdown-identification and -removal of social media content (see Kozyreva et al., 2023). Furthermore, as this background report will describe, prebunking interventions match a key part of the overall problem of misinformation: Most people only encounter misinformation occasionally and by accidental; would most often not pass it on; and mainly need of advice that allows them to avoid becoming confused or concerned as they read the misinformation.

What is misinformation?

Misinformation refers to any information that is false, incorrect, or misleading, irrespective of the intent behind its dissemination. It encompasses a broad spectrum of information that is not accurate, ranging from honest mistakes or misunderstandings to errors in data reporting or interpretation. Crucially, the key element of misinformation is that it is wrong or misleading, but it is not necessarily spread with a deliberate intention to deceive. This could occur when someone shares a rumor they believe to be true or a news story they have misinterpreted.

A related term is disinformation. Disinformation is a subset of misinformation that is deliberately created and spread with the intention to deceive or mislead. It is characterized by the purposeful production of false information or the manipulation of existing information to create a false narrative. The intent behind disinformation is often to cause harm, sow confusion, incite conflict, or influence public opinion or behavior in a certain way. Examples of disinformation include propaganda disseminated during political campaigns, hoaxes created to spread fear or panic, and manipulated images or videos used to misrepresent events or individuals.

The distinction between misinformation and disinformation is important from a government perspective as disinformation often requires more heavy-handed interventions including involvement of the police. From the perspective of the individual social media user the distinction is, however, less important. Whether or not a news story with false information was circulated deliberately or accidentally, the information contained in the news story is false and needs to be identified as such. Accordingly, we will here focus on the broader term of misinformation.

The prevalence of misinformation on social media

Challenges in estimating prevalence

It is difficult to make exact estimates about the prevalence of misinformation. One problem is to identify misinformation. Only a fraction of news stories or other types of information are fact-checked and, hence, it is very difficult to make estimates at the level of specific pieces of information or stories (i.e., how many percent of stories in current circulation on a social media platform are false?). Instead, researchers often rely on links to websites with a history of being untrustworthy and publishing false information. Yet sometimes these websites may publish true stories and sometimes completely mainstream media sources will publish stories that are misleading and even false.

The other problem is that 'prevalence' can be defined in multiple different ways: As the average number of pieces of misinformation shared on a social media platform relative to some metric for the total number of shared pieces of information; as the average number of users who share misinformation; and, finally, as the likelihood that a regular user will be exposed to misinformation. As we will show, these different quantities are not equally easy to estimate but, based on current knowledge, they provide completely different pictures of the magnitude of the problem of misinformation.

The final problem is that these estimates may vary significantly across platforms. We here focus on mainstream platforms such as Twitter and Facebook. Facebook is one of the most popular platforms of mainstream news consumption and Twitter plays a centrale role in information dissemination and media communication. Nonetheless, more research on the prevalence of misinformation on other popular platforms such as Tiktok, Youtube and Instagram is needed. Furthermore, some platforms, especially more politically extreme or fringe platforms, may have more significantly misinformation that what is discussed here.

Our current knowledge about the prevalence of misinformation on social media

Because of restrictions to data access on other platforms, the best available data is from Twitter and specifically focuses on the prevalence in social media posts of links to websites with a history of publishing false or untrustworthy information. This prevalence is often calculated as the percentage of links to such websites relative to the total number of links to news websites. Estimates here range from between 0.7 % to 6 % (Altay et al., 2022). These estimates are all from the United States, a highly polarized Western democracy. Note, however, that these estimates are relative to the total number of shared news links. If we compare to the actual number of posts, the estimate is much smaller. For example, one data set used to study misinformation contained 3,269 tweets with links to untrustworthy news sites, 85,344 tweets with links to a national news sites (trustworthy or not) and these tweets came from a large dataset that was comprised of 2.7 million tweets (Osmundsen et al., 2021). Only 0,001 % of the analyzed tweets thus contained links to a news site with a history of publishing false information.

Estimates of prevalence about misinformation on Facebook are substantially higher than Twitter estimates. Importantly, they also reflect something slightly different, namely user engagement with posts containing links to untrustworthy news sites with engagement defined as either clicks on the links or likes of the post. Estimates from United States suggest that these may be as high as 18-19 % of all engagements with news outlets on Facebook (Altay et al., 2021; Allen et al., 2021). Note, however, that this again is relative to other forms of news and not to the total number of posts on Facebook. Furthermore, estimates from other countries suggest significant variation, consistent with the marked role of national level factors in shaping vulnerability to disinformation and misinformation (Altay et al., 2021). Estimates are

higher (23 %) in France, a polarized European country, but much lower in the UK (2 %) and Germany (8.5 %) (Altay et al., 2021). These estimates are averages from 2017 to 2021.

Some studies on misinformation have looked at exposure to information from untrustworthy news sources (i.e., visits to them) rather than engagement with posts. These studies have focused on the US presidential elections and sought to estimate how many people were exposed to a single or more domains with a history of misinformation during the elections. Estimates show that 44 % of Americans were exposed in the 2016 election and 26 % in the 2020 election (Guess et al., 2019; Moore et al., 2023). These numbers are high but also suggest a substantial decrease in exposure to misinformation from one election to the other.

These estimates of prevalence range from low to high, depending on the specific metric, country, and platform. Yet, in all cases, it is key to note that these are averages. The problem with this is that averages are most informative in uniform distributions (i.e., distributions that are relatively centered around those averages) and when it comes to prevalence of misinformation the distributions are anything from uniform. This becomes clear when we shift our attention from prevalence at the post-level to prevalence at the user-level, i.e., change our question from how much misinformation is being shared on social media platforms to how many users share the misinformation?

Again, some of the best data comes from studies of Twitter in United States. Data here shows that only 1 % of users in a specific study was responsible for sharing a massive 75 % of all links to websites with a history of false news. 11 % of the users were responsible for sharing 100 % of such links (Osmundsen et al., 2021). The data from visits to untrustworthy websites during US elections also suggests that exposure to misinformation is concentrated to few individuals. Thus, while 26 % of Americans were exposed to one or more untrustworthy websites during the 2020 US presidential elections, the average number of visits was 23, suggesting that some visited a lot of websites and most just a single. Similarly, a narrower analysis of anti-vaccine content during the COVID-19 pandemic on Facebook and Twitter concludes that a massive 65 % of all anti-vaccine content was attributable to just twelve "super sharers" of misinformation (Nogara et al., 2022).

These highly concentrated numbers, in part, reflects that most people simply are not motivated to share any information online, trustworthy or untrustworthy. Thus, even for links to real news websites, 1 % of the users are responsible for sharing 30 % of all links (Osmundsen et al., 2021). People in general avoid active social media engagement especially on political topics (Osmundsen et al., 2021). The massive concentration of sharing among few individuals also reflects that the spread of content is shaped by network structure. Much misinformation is shared in highly structured networks, sometimes refers to as "echo chambers". Hence, while there is increasing agreement among researchers that online echo chambers are rare, they do exist and can be deep (Guess et al., 2018) and many of them are related to the extremist groups

associated with massive circulation of misinformation (e.g, Warner & Neville-Shepard, 2014). Among user centrally placed in such networks, circulation is substantial but likely much more infrequent outside.

A final complication is that not all misinformation is shared with the goal of misinforming. A study of Danish tweets about COVID-19 and facemasks sampled 9,345 tweets and found that 5 % of these engaged with misinformation either by tweeting or retweeting misinformation. Importantly, however, 50 % of those engagements were not actually supporting the misinformation but rather rejected it, in part by using humor to ridicule those who believed it (Johansen et al., 2022). As such, prevalence estimates for misinformation in regards to highly debated issues may be inflated by attempts to counter the misinformation.

Is the problem of misinformation getting worse?

Another way to assess the prevalence of individuals who are motivated to circulate misinformation is to assess historical trends. Are the number of individuals who could be motivated to share misinformation increasing or decreasing?

Assessing the spread of misinformation in a comparative way is difficult but there are a few studies that have tried to do just that. One study examined 120,000 letters to the editor in New York Times and Chicago Tribune from 1890 to 2010 and assessed these for conspiracy theories (Uscinski & Parent, 2014). The overall conclusion is that while the content and targets of conspiracy theories shift historically the volume of conspiracy theories in these outlets is highly stable historically. Another study sought to identify older surveys that examined Americans' beliefs in conspiracy theories and repeated those questions in 2020 (Uscinski et al., 2022). The oldest point of comparison was from 1966 and, hence, the study spanned more than half a century. Overall, the study concluded that there was no tendency for individuals who believe in conspiracy theories to become more prevalent. Finally, an already referenced study examined exposure to misinformation across four countries (US, UK, France and Germany) during the COVID-19 pandemic and in the years prior to the pandemic (Altay et al, 2022). Importantly, the study concluded that exposure to misinformation did not increase during the COVID-19 pandemic, despite the concerns regarding an 'infodemic' of misinformation.

This suggest again that the spread of misinformation on social media today is not a mass phenomenon, even if the prevalence of posts with misinformation is relatively high. It is confined to a small proportion of extremists who are relatively stable over longer periods of time. This should not be taken as an indication that the sharing of misinformation could not fluctuate historically and analyses suggests that misinformation does surge just prior to extremist events such as riots. Historical examples include ethnic riots (Horowitz) and a modern example is likely the storming of the US congress on January 6, 2021 (Arceneaux & Truex, 2022).).

Who spread and believe in misinformation?

Some psychological studies have looked at the underlying factors that drive misinformation sharing. Many of these studies rely on self-reported measures of intentions to share on social media a specific set of false news stories. Fewer studies rely on measures of actual sharing of false news stories on social media due to restrictions on data access and the logistical complications of combining behavioral data from social media with psychological measures obtained, for example, via surveys.

Cognitive and socio-affective factors underlying misinformation

Overall, these studies focus on two sets of factors: Cognitive factors and socio-affective factors (Ecker et al., 2022). Cognitive factors are factors related to the ability to discern the veracity of the information. A commonly studied factors is so-called cognitive reflection, i.e., the ability to override incorrect "gut" responses to a task and engage in further reflection to arrive at a correct answer (Pennycook & Rand, 2021. Cognitive reflection and related factors have been found to predict both beliefs in misinformation and, less consistently, the sharing of misinformation (Pennycook & Rand, 2021; Osmundsen et al., 2021).

Socio-affective factors are factors that influence the attraction or usefulness of a specific type of information, above and beyond its veracity. A standard view in social psychology is thus that people are motivated to believe in and promote information that aligns with their political and social motivation (Leeper & Slothuus, 2014). According to a number of studies one particularly strong motivation for sharing information relates to group-based conflict, whether these groups are political parties, nationalities or ethnic groups (e.g., Osmundsen et al., 2021; Horowitz, 2001; Mazepus et al., 2023). In such conflicts, people share information to mobilize their group against the perceived enemy and the better the information serves this purpose the more useful it is (Petersen, 2020). Misinformation, which often involves conspiracy theories or denigration of political groups, is often quite useful in this specific sense.

Consistent with the role of usefulness, both cross-national studies of self-reported motivations to share misinformation and studies of actual sharing of links to untrustworthy websites on US Twitter find that more partisan individuals, individuals who skepticial of the political system and authorities as well as individuals with more aggressive and conflict-oriented personalities are more likely to share (e.g., Mazepus et al., 2023; Osmundsen et al., 2021; Petersen et al., 2023). Importantly, these empirical relationships are not confined to misinformation per se. Partisan individuals, for example, are also more likely to share highly hostile abut true news, suggesting that the key factor is indeed the usefulness of the news stories for mobilization purposes and not their veracity (Osmundsen et al., 2021). In this sense, we can interpret the sharing of misinformation as an extreme form of political activism.

Socio-affective factors not only influence sharing decisions but also the propensity to believe in misinformation. People consistently are more likely to believe in information including conspiracy theories that align with their side in political and social conflicts. However, there is some evidence to suggest that socio-affective factors matter more for sharing decisions than belief formation whereas cognitive factors matter more for beliefs than sharing (Bor et al., 2023; Osmundsen et al., 2021; Pennycook & Rand, 2019). Indeed, people are also less likely to discern between false and true stories when making sharing decisions than when judging their accuracy (Arechar et al., 2022). This suggests that some stories are shared even when if they are viewed as inaccurate. Again, this aligns with the view that sharing is to a significant extent a form of political activism accomplished with specific goals in mind that extends beyond sharing accurate information.

Different individuals, different goals

These results suggest that the psychology governing sharing of both true and false information on social media is the same generic psychology that governs information sharing more generally and not something specific to social media contexts. This does not mean that social media does not make a difference for the effectiveness with which people can satisfy these more generic cognitive and social goals.

Social media provides individuals with greater connectedness. This connectedness gives people a potential longer reach, greater opportunities for acting fast and a better possibility of finding like-minded others. In this perspective, social media does not change the underlying psychological goals of information-sharing but is a tool affecting the effectiveness with which people can meet these goals (Bor & Petersen, 2022).

It is important to note that the same goals are not likely to be equally strongly present in everyone. Some people's sharing decisions are likely more strongly governed by cognitive goals and some people's sharing decisions are likely governed more by socio-affective factors. For example, research suggest that individuals with a strong emotional attachment to a political party are more likely to feel socio-affective motivations to defend and advance the standing of this party (Huddy et al., 2015) through, for example, the sharing of highly partisan or false information (Osmundsen et al., 2021). Similarly, other research show that the mindset to knowingly share misinformation is rooted in extreme feelings of loneliness and marginalization (Petersen et al., 2023). As such, it is possible that some cognitively-motivated individuals accidentally share a few pieces of misinformation because they fail to identify them as false, whereas the habitual sharers responsible for the bulk of misinformation are motivated by political goals of a socio-affective nature and, hence, focus much more on the usefulness of the information than its veracity.

It is important to add two points of clarification with respect to the role of veracity in habitual sharers decision-making process. First, even habitual sharers may not actively process the stories they share as false before they share. It is likely that considerations about veracity does not enter strongly into their decision-making process when deciding to share or not. Second, for habitual sharers motivated by political-activist goals, many of pieces of misinformation could seem true, even if they did more strongly consider this dimension. Those who are motivated to share misinformation often operate on the basis of a worldview that is characterized by deep mistrust in official sources and mainstream institutions (Petersen et al., 2023). On the background of such a worldview, it may seem plausible that political elites and other mainstream actors could engage in the preposterous actions that conspiracy theories accuse them of.

The effects of exposure to misinformation

A key concern in public debates about the online spread of misinformation is the extent to which the misinformation impacts the attitudes and beliefs of those exposed to it. In general research has provided significant insights into the general persuasive effects of exposure to information through laboratory and survey experiments (Druckman, 2022). However, there are practical and ethical problems related to exposing people to misinformation and, accordingly, there are few experimental studies of the effect of misinformation. On the basis of our general understanding of persuasion effect as well as some studies about misinformation specifically with non-experimental research designs, it is, however, possible to pin together a well-founded understanding of the effects of being exposed to misinformation.

Overall, it is likely that the effect of any single piece of misinformation is very small (Altay et al., 2023), especially in countries with well-functioning media institutions (Humprecht et al., 2020). The circulation of misinformation can, however, have other detrimental effects.

Research suggests that there is a clear association between exposure to misinformation and belief in misinformation (Roozenbeek et al., 2020). However, correlation is not causation, and the association does not imply that exposure to misinformation is the main factor in explaining misinformed beliefs or behavior (de Saint Laurent et al., 2022). Specifically, we know that information access – including access to partisan information in online environments (Robertson et al., 2023) – is heavily shaped by so-called selective exposure, that is, that people select information that aligns with their pre-existing beliefs and attitudes (Smith et al., 2008). Accordingly, the underlying worldview of distrusting political and media institutions is often more a cause rather an effect of misinformation exposure, for example, by motivating people to select into the social media networks where such misinformation is circulated (Petersen et al., 2023).

Nonetheless, the circulation of misinformation may have two distinct effects on those already holding polarized and conspiratorial worldviews. First, even if concrete pieces of misinformation do not cause habitual sharers' worldviews, the concrete pieces of misinformation can help them justify and defend these worldviews (Williams, 2023). Second, the circulation of misinformation on social media can help coordinate the political activism of likeminded individuals (Petersen, 2020). An example in point is the storm on the US congress in January 2021. While the circulation of misinformation regarding electoral fraud was likely more an effect rather than a cause of the underlying worldviews, dispositions and frustrations of those who believed in misinformation about election fraud (Arceneaux & Truex, 2022), this circulation nonetheless facilitated the mobilization at the particular time and place with the particular goal in mind of storming the US Congress.

Even if the circulation of misinformation is often more an effect rather than a cause of habitual sharers' worldviews, a key question is still is how exposure to misinformation impacts those who are more incidentally exposed to misinformation and who do not share the underlying worldviews promoted in conspiracy theories and partisan worldviews. Here, research suggest that their pre-existing – i.e., less extremist, partisan and distrusting – attitudes serve as a bulwark against misinformation. People are most often not strongly swayed by any information (Mercier, 2017) and, in particular, not by information that does not align with what they already believe (Slothuus, 2010). For example, a massive study on the effects of 49 field experiments on political campaigns suggest that "the effects of campaign contact and advertising on Americans' candidates choices in general elections is zero" (Kalla & Broockman, 2017). Another massive study examined the effect of a 8-month-long advertising programme delivered via a social media to persuade 2 million voters to vote for Biden rather than Trump in the 2020 US presidential election (Aggarwal et al., 2023). The cost of this programme was $8.9 million and it increased turnout among Biden leaners with 0.4 percentage points and decreased turnout among Trump leaners with 0.3 percentage points, leading the authors to conclude that "effects of even large digital advertising campaigns in presidential elections are likely to be modest."

These findings on the general persuasiveness of information suggests that the effects of misinformation is small as well. A direct demonstration of this point is the research literature on the persuasiveness of some of the history's most nefarious misinformation, the propaganda in Nazi Germany. The common finding in this literature is that speeches of Adolf Hitler had surprisingly little effect on support for the Nazi party and mostly had effects on those already disposed to support the party (e.g., Selb & Munzert, 2018; Adena et al., 2015). Indoctrination, instead, occurred primarily through the school system rather than mass media (Voigtländer & Voth, 2015) A more recent demonstration focused on the effects of exposure to malicious Russian social media accounts among American users during the 2016 presidential election and found no effect (Eady et al., 2023).

At the same time, however, this does not necessarily entail that the effects of exposure to misinformation among those incidentally exposed is always zero. One study conducted early in the pandemic of exposure to misinformation regarding COVID-19 vaccinations finds a significant effect of exposure to misinformation on COVID-19 vaccination intentions among Americans and British people (Loomba et al., 2021). Specifically, exposure reduces the share of individuals who would "definitely" take the vaccine with 6.2 and 6.4 percentage points, respective. As a general indication of the persuasiveness of misinformation, this estimate could potentially be seen as an estimate of the upper-bound on persuasiveness, at least, within Western democracies. The study was conducted prior to the availability of the vaccines and the the COVID-19 pandemic characterized by an unusual degree of uncertainty and shifting information. Furthermore, researchers argue that the resistance to misinformation varies significantly across countries with the US being one of the countries with least resistance (Humprecht et al., 2020). People in countries with well-functioning and trusted mainstream media institutions as well as a lack of polarization are much more resistant to misinformation as the media and authorities can continuously set the record straight on those pieces of misinformation stories that receive significant attention.

While the direct effects of misinformation on attitudes and beliefs is likely very small, the intense circulation may have other detrimental consequences. Misinformation can distract people's time and attention from more relevant information sources; misinformation can deplete authorities' and mainstream media's resources because of a constant felt need to counter it; and, even if misinformation does not change beliefs and attitudes, it can still sow confusion and uncertainty until the information is identified as false.

A problem-diagnosis: The two problems of misinformation

This literature review about the prevalence, causes and effects of misinformation suggests that it is unhelpful to talk about the problem of misinformation in singular. It is, in fact, two problems and these are highly very different.

First, there is the problem of habitual sharers. These few individuals are responsible for the sharing of the vast majority of misinformation in relatively closed networks – as well as the sharing of a lot of true information. They are motivated by political goals and share information that is aligned with those goals. These political motivations are often rooted in larger frustrations with society or societal issues and, hence, are difficult to change without the implementation of actual policies that address the issues and their underlying drivers such as marginalization or polarization. In the case of the more extremist individuals, this may also require genuine deradicalization interventions. As should be clear, it is not easy to solve the problem of habitual sharers without continued political effort as the magnitude of the problem to a large extent will reflect long-term societal conflicts. In this perspective, the amount of

misinformation circulated by habitual sharers is in many ways a "canary in a coal mine", indicating the magnitude of wider societal issues.

Second, there is the problem of incidental exposure. The few habitual sharers are the drivers of the spread of misinformation and most people meet misinformation incidentally through exposure to the activities of habitual sharers as these activities occasionally spill-over from more closed online networks to the larger mainstream network. This majority is likely to be motivated to believe in and share accurate information (Pennycook et al., 2021). Upon exposure, these individuals need to be able to discern between true and false information in order to avoid uncertainty and confusion and, even if rare, accepting the false information as true.

These two problems differ dramatically. The problem of habitual sharers is the most concerning problem as these habitual sharers also pursue their political goals through other means including by support the use of political violence (Petersen et al., 2023). Yet, this problem may not be solvable using scalable online interventions. In contrast, the problem of incidental exposure is the more common problem and the one may be solvable using scalable online interventions.

Remedying the problem of the more common problem of incidental exposure involves a conceptual shift from focusing interventions on changing the motivations of habitual sharers to mitigating the effects of their actions on the beliefs of naive bystanders. Importantly, empowering those who are incidentally exposed to misinformation to discern between true and false is a much less formidable, even if not easy, task than changing the activist mindsets of the habitual sharers. This task is the focus on the reminding sections of this report.

Countering the adverse consequence of incidental exposure to misinformation

A key issue facing any intervention designed to empower users to identify true from false is the scalability of the intervention, i.e., the ability to deploy the intervention to a sufficiently larger number of users. Currently, a number of scalable interventions exists and these fall generally into two large categories: Interventions designed to *debunk* and interventions designed to *prebunk* misinformation.

Debunking via fact-checking

Debunking is the general term for interventions that identify and inform users that particular news stories or information are inaccurate. Thus, debunking is interventions focused on mitigating the effects of misinformation that is already spreading. The most common type of

debunking occurs in collaboration between fact-checkers and social media companies where journalists or other fact-checkers identifies that a circulating piece of information is false or misleading and social media companies subsequently flags or labels this piece of information as misleading. Research shows that such fact-checks works but also that the effects are often small and reminders are necessary (Carey et al., 2022). Importantly, as argued above, a small and effect of fact-checks will not necessarily reflect that people do not care about truth but rather that most people most of the time are already quite good at discerning between true and false information and seldomly engages in the sharing of any information. Furthermore, in addition to any effects on citizens' abilities to avoid sharing misinformation, fact-checking institutions have other important functions on the overall quality of democratic conversations. Research thus shows that fact-checking have effects on political elites such that politicians being fact-checked are subsequently less likely to share misleading information (Lim, 2018; Nyhan & Reifler, 2015).

Debunking interventions are valuable tools when single pieces of misinformation get wide traction and to ensure that political elites are accountable. At the same time, debunking interventions faces two challenges. First, they are fully contingent on public trust in the debunkers such as fact-checking institutions and the media that disseminations the fact-checks to the wider public. As such that the probably most effective in the high-trust countries that are already least vulnerable to misinformation (Humprecht et al., 2018). Second, debunking takes time and costs money. It is simply impossible to debunk all pieces of misinformation that are circulating at any given point in time. As such, innocent bystanders could become accidental victims of pieces of misinformation that are not getting sufficient widespread attention to warrant fact-checking. The other type of interventions, prebunking interventions, are designed to remedy both these weaknesses.

Prebunking interventions

Prebunking to the proactive process of debunking false information before it has had a chance to spread widely or before a person has been exposed to it (Lewandowsky & Van Der Linden, 2021). Prebunking aims to create a form of "cognitive immunity" against misinformation in general, so that when people encounter any piece of misinformation, they are better prepared to critically evaluate information and less likely to be misled. In that sense, prebunking interventions are preventative, as opposed to debunking, which is reactive and is oriented towards a specific piece of misinformation that has already been circulated.

While the concept of debunking is closely tied to a single form of intervention (i.e., fact-checking), the concept of prebunking is tied to a wider range of specific interventions. Some of these interventions increases the motivation to be vigilant against misinformation. Other interventions increase the ability to engage in vigilance with success.

Accuracy nudges

The most prominent motivation-oriented prebunking intervention is the so-called accuracy nudge (Pennycook et al., 2021). This intervention is premised on the argument that most people are motivated to share accurate information but occasionally are distracted by the busyness of social media feeds to not invest sufficiently in telling true from false. Accuracy nudges are gentle reminders that not all information users meet is accurate. The primary advantage of accuracy nudges are that they are highly scalable and can easily be implemented by social media platforms. The overall empirical evidence suggest that gentle accuracy nudges do increase the relative sharing of true over false news stories both in laboratory experiments and on actual social media platforms (Pennycook & Rand, 2022). At the same time, critical examinations have noted that the effects of such nudges on are small (e.g., Rathje et al., 2022; Rasmussen et al., 2022). This could be seen as consistent with the finding that there is already widespread public concern over misinformation and "fake news." The problem may, in other words, not be a lack of motivation but a lack of competence in discerning between true and false among those who incidentally pass on misleading misinformation. Remedying this problem is the focus of two forms of competence-oriented prebunking interventions.

Gamified innoculation

Some competence-oriented prebunking interventions are focused on so-called inoculation and often utilizes gamification approaches. The theoretical idea of inoculation comes from vaccination. Like with a vaccine, the argument goes, a limited and controlled degree of exposure to misinformation will immunize the receiver against subsequent exposures (Lewandowsky & Van Der Linden, 2021).

While the vaccination analogy may not be perfect, there is substantial evidence that people can build competences to detect misinformation through gamified interventions (Roozenbeek et al., 2020). The immediate effects of innoculation are medium large in size (Traberg et al., 2022); the effects are detectable over periods of several weeks (one study demonstrated that the effect was detectable over a 13-week period) (Maertens et al., 2021); and have been observed both in the laboratory and actual social media platforms (Roozenbeek et al., 2022).

Methodologically, many of these interventions have utilized a gamified approach. Most prominently, the game "Bad News" has been designed to provide players with an understanding of the format and content of viral false news stories (https://www.getbadnews.com/books/english/). Specifically, in this game, players construct false news stories out of a range of available building blocks. They subsequently score points accordingly to how well their news stories fit an evidence-based template of viral misinformation. Many other games of a similar kind have been developed by the researchers such as the Go Viral game (https://www.goviralgame.com/books/go-viral/), designed for

countering COVID-19 related misinformation during the pandemic, or the Cranky Uncle game (https://crankyuncle.com/game/), designed for countering misinformation regarding climate change. These games have been tested in scientific studies and vary in time between 6 minutes to 15 minutes. The accuracy nudge reviewed above, in contrast, takes maybe 30 seconds to complete. As such, gamified interventions are by nature relatively long and, hence, less scalable, even if they appear more effective.

Prebunking by increasing digital literacy

A more traditional competence-oriented prebunking interventions is designed to increase the digital literacy of receiver. Digital literacy is a broad concept that is often defined as ability to understand and navigate digital environments including by searching for and identifying valuable information (sometimes referred more specifically as information literacy; Bawden, 2008).

In the context of misinformation, digital literacy interventions often involve short (i.e., processable in a few minutes) educational videos or texts that provide the receiver with advice on how to identify misinformation on social media, essentially aimed at turning the user into their own fact-checker. For example, in a number of countries, Facebook circulated a set of "tips to spot false news" (Guess et al., 2020). The verbatim content of these tips was:

- **Be skeptical of headlines.** False news stories often have catchy headlines in all caps with exclamation points. If shocking claims in the headline sound unbelievable, they probably are.
- **Look closely at the URL.** A phony or look-alike URL may be a warning sign of false news. Many false news sites mimic authentic news sources by making small changes to the URL. You can go to the site to compare the URL to established sources.
- **Investigate the source.** Ensure that the story is written by a source that you trust with a reputation for accuracy. If the story comes from an unfamiliar organization, check their "About" section to learn more.
- **Watch for unusual formatting.** Many false news sites have misspellings or awkward layouts. Read carefully if you see these signs.
- **Consider the photos.** False news stories often contain manipulated images or videos. Sometimes the photo may be authentic, but taken out of context. You can search for the photo or image to verify where it came from.
- **Inspect the dates.** False news stories may contain timelines that make no sense, or event dates that have been altered.
- **Check the evidence.** Check the author's sources to confirm that they are accurate. Lack of evidence or reliance on unnamed experts may indicate a false news story. Look at other reports. If no other news source is reporting the same story, it may indicate that the story is false. If the story is reported by multiple sources you trust, it's more likely to be true.

- **Is the story a joke?** Sometimes false news stories can be hard to distinguish from humor or satire. Check whether the source is known for parody, and whether the story's details and tone suggest it may be just for fun. Some stories are intentionally false. Think critically about the stories you read, and only share news that you know to be credible.

Versions of these tips appear in multiple digital literacy interventions. For example, Table 1 includes a list of evidence-based educational videos and abbreviated versions of their content, evaluated for effectiveness by Bor et al. (2003).

Research has demonstrated that these digital media literacy interventions do in fact increase people's ability to discern between true and false news upon exposure. Guess et al. (2020) concludes that the Facebook digital literacy intervention "has the largest measured effect size to date on ratings of false headlines". This effect is replicated, albeit with a smaller effect, in India. Similarly, Bor et al. (2023) also finds small to medium-sized effects of the educational videos listed in Table 1 on abilities to discern true from false news stories. At the same time, it is worth noting that, as with fact-checks, the effect of digital literacy interventions seems to decay over time (Guess et al., 2020). Accordingly, such interventions need to be a routine part of social media environments and they need to be sufficiently engaging for people to repeatedly watch different versions of them.

Furthermore, it is important to note that digital media literacy may not impact the sharing of misinformation. As discussed above, those who routinely share misinformation are most often tech-savy activists and, hence, often they are quite high in digital media literacy (Osmundsen et al., 2021; Sirlin et al., 2021). Most people without digital media literacy would not share any information – true or false – on social media in the first place. The key contribution is thus whether prebunking interventions will protect these individuals from believing in the false information that the accidentally and occasionally encounter. The current evidence suggests that this is the case.

The relevance of literacy-oriented prebunking is also consistent with the broader research literature on risk communication and behavior. Decades of research in risk communication demonstrates that the key way to facilitate behavior change in the face of risk is to build so-called coping ability, i.e., build a sense of efficacy and competence through concrete actionable advice (Norman et al., 2015). This sense of coping ability has been found to be more important in shaping behavior change towards threats compared to feelings of anxiety or concern regarding the threat. Knowing what to do and how to do it is key for adaptive behavior, whereas anxiety may generate action impulses that leads to inefficient behavior. Recent research suggest that digital literacy interventions do indeed shape feelings

of efficacy and that these feelings of efficacy are closely related to the actual ability to discern between valid information and misinformation (Rasmussen et al., 2022).

If coping ability is established through an intervention, including information about the risks involved in the threat will often (but not always; Jørgensen et al., 2021) increase motivations to engage in protective behavior against. As such, a particularly strong prebunking intervention against misinformation may combine both motivation- and competence-oriented information. Creating such stronger interventions may be especially important as artificial intelligence makes it possible to create more realistic forms of misinformation (including false videos or so-called deep fakes). AI-powered misinformation will likely be harder to detect and therefore require the additional effort motivated by a stronger intervention (Vaccari & Chadwick 2020).

The danger of skepticism and the importance of humility

The focus of most interventions against misinformation is to activate informational vigilance among citizens. A central concern in the associated research literature has been whether the vigilance caused by interventions exclusively decrease belief in and motivations to share false news; or, alternatively, whether these interventions activate a general mistrust of the information encountered on the Internet such that people also decrease their beliefs in and circulation of real and valid news. In the latter case, the ratio of real to false information in circulation may not change much because of the intervention.

To counter against this latter possibility, many studies of misinformation interventions focus on the effects of those interventions on so-called discernment, i.e., the relative effect on beliefs in true versus false news. The prebunking interventions discussed above have thus all be found to shape discernment. Nonetheless, the concern is real. Other research suggests that such interventions also increase skepticism about accurate news (Guess et al., 2020; Hoes et al., 2023); that interventions warning about misinformation can negatively affect trust in science and politics (Hoes et al., 2022); and that there is a general association between perceiving misinformation as widespread and distrusting the media (Hameleers et al., 2022).

More generally, distrust and skepticism have been found to correlate positively with the tendency to share and believe in misinformation and conspiracy theories (Petersen et al., 2023). Conspiracy theorists in general also see themselves as critical, science-minded and evidence-based (Marie & Petersen, 2022). Instead, there is emerging evidence that a key psychological trait that predicts lack of beliefs and lack of motivation to share false or hyperpartisan information is trait of "intellectual humility" (Marie & Petersen, 2022). Intellectual humility "involves being humble with regard to the way one acquires and applies knowledge" (Krumrei-Mancuso & Rouse, 2016) and, hence, is the exact opposite of the moralized sense of being highly knowledgable that is associated with conspiracy beliefs (Marie & Petersen, 2022). From

an interventionist perspective, it is also important to note that research shows that it is in fact possible to activate a sense of humility by showing people that their intuitions are fallible (Koetke et al., 2023).

These findings serve as an important warning that interventions should not simply focus on activating psychological states of suspicion, skepticism, and critique. Such interventions may backfire, at least, if they do not also teach actual tools of analytical thinking. The underlying tone of any intervention should instead resonate with the goal of inducing a state of humility. This is consistent with the basic findings of the drivers of belief in misinformation. Humility is required to work against the influence of socio-affective biases and engage in the hard cognitive work of reasoning to discern between true and false.

Table 1. Overview of Six Video-Based Educational Digital Media Literacy Interventions

Title	Author	Link	Number of views	Advice (abbreviated form)
How to choose your news - Damon Brown	Ted ED	youtu.be/q-Y-z6HmRgI	1,100,000	1. Get original news cutting out the middlemen 2. In chaotic times, do not follow news constantly 3. Check multiple sources 4. Be aware of anonymous sources 5. Verify before spreading
Five ways to spot fake news	Quartz	youtu.be/y7eCB2F89K8	150,000	1. Consider where the information is coming from 2. Consider if the headline sound neutral 3. Consider who wrote it 4. Consider what the resources are 5. Consider if the images are accurate
How to Spot Fake News	Factcheck.org	youtu.be/AkwWcHekMdo	550,000	1. Consider the source 2. Read beyond headline 3. Check the author 4. Consider what the support is for a claim 5. Check the date 6. Consider if it could be satire 7. Check biases 8. Consult experts
The Fact Checker's guide for detecting fake news	Washington Post	youtu.be/SoWCDJAMk2Y	26,000	1. Double check the url 2. Consider if the photo seems unrealistic

				3. Check the sources 4. Use dedicated plug-ins that can connect to fact-checkers
Helping Students Identify Fake News with the Five C's of Critical Consuming	John Spencer	youtu.be/xf8mjbVRqao	363,000	1. Context 2. Credibility 3. Construction 4. Corroboration 5. Compare
Four ways to tell if something is true online - Break the Fake	MediaSmarts	youtu.be/E-049KTrYBg	10,000	1. Use fact checking tools 2. Find the source 3. Verify the source 4. Check other sources

Notes. Copied from Bor et al. (2023). The number of views is per December 6 2022.

References

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., & Zhuravskaya, E. (2015). Radio and the Rise of the Nazis in Prewar Germany. *The Quarterly Journal of Economics*, *130*(4), 1885-1939.

Aggarwal, M., Allen, J., Coppock, A., Frankowski, D., Messing, S., Zhang, K., ... & Zheng, S. (2023). A 2 million-person, campaign-wide field experiment shows how digital advertising affects voter turnout. *Nature Human Behaviour*, 1-10.

Allen, J., Mobius, M., Rothschild, D. M., & Watts, D. J. (2021). Research note: Examining potential bias in large-scale censored data. *Harvard Kennedy School Misinformation Review*.

Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on misinformation: Conceptual and methodological challenges. *Social Media+ Society*, *9*(1), 20563051221150412.

Altay, S., Nielsen, R. K., & Fletcher, R. (2022). Quantifying the "infodemic": People turned to trustworthy news outlets during the 2020 coronavirus pandemic. *Journal of Quantitative Description: Digital Media*, *2*.

Arceneaux, K., & Truex, R. (2022). Donald Trump and the lie. *Perspectives on Politics*, 1-17.

Arechar, A. A., Allen, J. N. L., Cole, R., Epstein, Z., Garimella, K., Gully, A., ... & Rand, D. (2022). Understanding and reducing online misinformation across 16 countries on six continents. PsyArXiv. February 11. doi:10.31234/osf.io/a9frz.

Bawden, D. (2008). Origins and concepts of digital literacy. *Digital literacies: Concepts, policies and practices*, *30*(2008), 17-32.

Bor, A., & Petersen, M. B. (2022). The psychology of online political hostility: A comprehensive, cross-national test of the mismatch hypothesis. *American political science review*, *116*(1), 1-18.

Bor, A., Osmundsen, M., Rasmussen, S. H. R., Bechmann, A., & Petersen, M. B. (2023). "Fact-checking" videos reduce belief in misinformation and improve the quality of news shared on Twitter. PsyArXiv. September 24. doi:10.31234/osf.io/a7huq

Briand, S. C., Cinelli, M., Nguyen, T., Lewis, R., Prybylski, D., Valensise, C. M., ... & Quattrociocchi, W. (2021). Infodemics: A new challenge for public health. *Cell*, *184*(25), 6010-6014.

Carey, J. M., Guess, A. M., Loewen, P. J., Merkley, E., Nyhan, B., Phillips, J. B., & Reifler, J. (2022). The ephemeral effects of fact-checks on COVID-19 misperceptions in the United States, Great Britain and Canada. *Nature Human Behaviour*, *6*(2), 236-243.

de Saint Laurent, C., Murphy, G., Hegarty, K., & Greene, C. M. (2022). Measuring the effects of misinformation exposure and beliefs on behavioural intentions: A COVID-19 vaccination study. *Cognitive Research: Principles and Implications*, *7*(1), 87.

Druckman, J. N. (2022). A Framework for the Study of Persuasion. *Annual Review of Political Science*, *25*, 65-88.

Eady, G., Paskhalis, T., Zilinsky, J., Bonneau, R., Nagler, J., & Tucker, J. A. (2023). Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior. *Nature Communications*, *14*(1), 62.

Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., ... & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), 13-29.

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

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.

Guess, A., Nyhan, B., Lyons, B., & Reifler, J. (2018). Avoiding the echo chamber about echo chambers. *Knight Foundation*, *2*(1), 1-25.

Hameleers, M., Brosius, A., & de Vreese, C. H. (2022). Whom to trust? Media exposure patterns of citizens with perceptions of misinformation and disinformation related to the news media. *European Journal of Communication*, *37*(3), 237-268.

Hoes, E., Aitken, B., Zhang, J., Gackowski, T., & Wojcieszak, M. (2023). Prominent Misinformation Interventions Reduce Misperceptions but Increase Skepticism. PsyArXiv. May 10. doi:10.31234/osf.io/zmpdu.

Hoes, E., Clemm, B., Gessler, T., Wojcieszak, M., & Qian, S. (2022). The cure worse than the disease? PsyArXiv. November 12. doi:10.31234/osf.io/4m92p.

Horowitz, D. L. (2001). *The deadly ethnic riot*. Univ of California Press.

Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *American Political Science Review*, *109*(1), 1-17.

Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to online disinformation: A framework for cross-national comparative research. *The International Journal of Press/Politics*, *25*(3), 493-516.

Johansen, N., Marjanovic, S. V., Kjaer, C. V., Baglini, R. B., & Adler-Nissen, R. (2022). Ridiculing the "tinfoil hats:" Citizen responses to COVID-19 misinformation in the Danish facemask debate on Twitter.". *Harvard Misinformation Review*, *3*(2).

Jørgensen, F., Bor, A., & Petersen, M. B. (2021). Compliance without fear: Individual-level protective behaviour during the first wave of the COVID-19 pandemic. *British Journal of Health Psychology*, *26*(2), 679-696.

Kalla, J. L., & Broockman, D. E. (2018). The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments. *American Political Science Review*, *112*(1), 148-166.

Koetke, J., Schumann, K., Porter, T., & Smilo-Morgan, I. (2023). Fallibility salience increases intellectual humility: Implications for people's willingness to investigate political misinformation. *Personality and Social Psychology Bulletin*, *49*(5), 806-820.

Kozyreva, A., Herzog, S. M., Lewandowsky, S., Hertwig, R., Lorenz-Spreen, P., Leiser, M., & Reifler, J. (2023). Resolving content moderation dilemmas between free speech and harmful misinformation. *Proceedings of the National Academy of Sciences*, *120*(7), e2210666120.

Krumrei-Mancuso, E. J., & Rouse, S. V. (2016). The development and validation of the comprehensive intellectual humility scale. *Journal of Personality Assessment*, *98*(2), 209-221.

Leeper, T. J., & Slothuus, R. (2014). Political parties, motivated reasoning, and public opinion formation. *Political Psychology*, *35*, 129-156.

Lewandowsky, S., & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, *32*(2), 348-384.

Lim, C. (2018) Can Fact-checking Prevent Politicians from Lying? Preprint, https://discuss.tp4.ir/uploads/default/original/2X/6/620e0f36b3d2898e3a4672aa572cb0c950448ed0.pdf

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature human behaviour*, *5*(3), 337-348.

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

Marie, A., & Petersen, M. B. (2022) Moralization of rationality can stimulate sharing of hostile political news, but intellectual humility inhibits it. OSF Preprints. March 4. doi:10.31219/osf.io/k7u68.

Mazepus, H., Osmudsen, M., Bang-Petersen, M., Toshkov, D., & Dimitrova, A. (2023). Information battleground: Conflict perceptions motivate the belief in and sharing of misinformation about the adversary. *Plos one*, *18*(3), e0282308.

Mercier, H. (2017). How gullible are we? A review of the evidence from psychology and social science. *Review of General Psychology*, *21*(2), 103-122.

Moore, R. C., Dahlke, R., & Hancock, J. T. (2023). Exposure to untrustworthy websites in the 2020 US election. *Nature Human Behaviour*, 1-10.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D., & Nielsen, R. K. (2018). Reuters institute digital news report 2018. *Report of the Reuters Institute for the Study of Journalism*.

Nogara, G., Vishnuprasad, P. S., Cardoso, F., Ayoub, O., Giordano, S., & Luceri, L. (2022, June). The disinformation dozen: An exploratory analysis of covid-19 disinformation proliferation on twitter. In *14th ACM Web Science Conference 2022* (pp. 348-358).

Norman, P., Boer, H., Seydel, E. R., & Mullan, B. (2015). Protection motivation theory. *Predicting and changing health behaviour: Research and practice with social cognition models*, *3*, 70-106

Nyhan, B., & Reifler, J. (2015). The effect of fact-checking on elites: A field experiment on US state legislators. *American Journal of Political Science*, *59*(3), 628-640.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. *American Political Science Review*, *115*(3), 999-1015.

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.

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in cognitive sciences*, *25*(5), 388-402.

Pennycook, G., & Rand, D. G. (2022). Nudging social media toward accuracy. *The Annals of the American Academy of Political and Social Science*, *700*(1), 152-164.

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590-595.

Petersen, M. B. (2020). The evolutionary psychology of mass mobilization: how disinformation and demagogues coordinate rather than manipulate. *Current opinion in psychology*, *35*, 71-75.

Petersen, M. B., Osmundsen, M., & Arceneaux, K. (2023). The "need for chaos" and motivations to share hostile political rumors. *American Political Science Review,* 1-20. doi:10.1017/S0003055422001447

Rasmussen, J., Lindekilde, L., & Petersen, M. B. (2022). Public health communication decreases false headline sharing by boosting self-efficacy. PsyArXiv. July 7. doi:10.31234/osf.io/8wdfp.

Rathje, S., Roozenbeek, J., Traberg, C. S., Van Bavel, J. J., & van der Linden, S. (2022). Letter to the editors of Psychological Science: meta-analysis reveals that accuracy nudges have little to no effect for US conservatives: regarding Pennycook et al.(2020).

Robertson, R. E., Green, J., Ruck, D. J., Ognyanova, K., Wilson, C., & Lazer, D. (2023). Users choose to engage with more partisan news than they are exposed to on Google Search. *Nature*, 1-7.

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., ... & Van Der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society open science*, *7*(10), 201199.

Roozenbeek, J., van der Linden, S., & Nygren, T. (2020). Prebunking interventions based on ''inoculation''theory can reduce susceptibility to misinformation across cultures. Harvard Kennedy School (HKS) Misinformation Review, 1 (2).

Roozenbeek, J., Van Der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science advances*, *8*(34), eabo6254.

Selb, P., & Munzert, S. (2018). Examining a most likely case for strong campaign effects: Hitler's speeches and the rise of the Nazi party, 1927–1933. *American Political Science Review*, *112*(4), 1050-1066.

Sirlin, N., Epstein, Z., Arechar, A. A., & Rand, D. G. (2021). Digital literacy is associated with more discerning accuracy judgments but not sharing intentions.

Slothuus, R. (2010). When can political parties lead public opinion? Evidence from a natural experiment. *Political Communication*, *27*(2), 158-177.

Smith, S. M., Fabrigar, L. R., & Norris, M. E. (2008). Reflecting on six decades of selective exposure research: Progress, challenges, and opportunities. *Social and Personality Psychology Compass*, *2*(1), 464-493.

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

Uscinski, J. E., & Parent, J. M. (2014). *American conspiracy theories*. Oxford University Press.

Uscinski, J., Enders, A., Klofstad, C., Seelig, M., Drochon, H., Premaratne, K., & Murthi, M. (2022). Have beliefs in conspiracy theories increased over time?. *PLoS One*, *17*(7), e0270429.

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media+ Society*, *6*(1), 2056305120903408.

Voigtländer, N., & Voth, H. J. (2015). Nazi indoctrination and anti-Semitic beliefs in Germany. *Proceedings of the National Academy of Sciences*, *112*(26), 7931-7936.

Warner, B. R., & Neville-Shepard, R. (2014). Echoes of a conspiracy: Birthers, truthers, and the cultivation of extremism. *Communication Quarterly*, *62*(1), 1-17.

Williams, D. (2023). The marketplace of rationalizations. *Economics & Philosophy*, *39*(1), 99-123.