



Negative Downstream Effects of Alarmist Disinformation Discourse: Evidence from the United States

Andreas Jungherr¹ · Adrian Rauchfleisch²

Accepted: 1 January 2024 / Published online: 12 January 2024
© The Author(s) 2024

Abstract

The threat of disinformation features strongly in public discourse, but scientific findings remain conflicted about disinformation effects and reach. Accordingly, indiscriminate warnings about disinformation risk overestimating its effects and associated dangers. Balanced accounts that document the presence of digital disinformation while accounting for empirically established limits offer a promising alternative. In a preregistered experiment, U.S. respondents were exposed to two treatments designed to resemble typical journalistic contributions discussing disinformation. The treatment emphasizing the dangers of disinformation indiscriminately (T1) raised the perceived dangers of disinformation among recipients. The balanced treatment (T2) lowered the perceived threat level. T1, but not T2, had negative downstream effects, increasing respondent support for heavily restrictive regulation of speech in digital communication environments. Overall, we see a positive correlation among all respondents between the perceived threat of disinformation to societies and dissatisfaction with the current state of democracy.

Keywords Disinformation · Discourse · Regulation · Democracy · Media effects · Experiment

✉ Andreas Jungherr
andreas.jungherr@uni-bamberg.com

¹ Institute for Political Science, University of Bamberg, Bamberg, Germany

² Graduate Institute of Journalism, National Taiwan University, Taipei, Taiwan

Negative Downstream Effects of Alarmist Disinformation Discourse

The threat of disinformation features strongly in discourse (Carlson, 2020; Gutsche, 2018).¹ Many characterize digital communication environments as especially vulnerable to intentional disinformation or accidental misinformation.² Perceived drivers include heterogeneity of information sources, weakened power of gatekeepers in establishing and enforcing information quality standards and processes, and information flows shaped by algorithms or targeted interventions through ads. In this discourse, people are generally portrayed as being easily manipulated through false or misleading information (Gutsche, 2018). General warnings that digital information environments are unreliable and spaces of broad society-wide manipulation abound in journalistic accounts; communiqués by governing bodies, such as from the EU (High Representative of the Union for Foreign Affairs and Security Policy, 2018); court rulings, such as a decision by Germany’s supreme court (BVerfG, 2021); and from politicians who attack companies that provide digital infrastructures, such as U.S. President Joe Biden denouncing Facebook for allowing the spread of disinformation on Covid: “They’re killing people” (Kanno-Young & Kang, 2021).

Although selected scientific findings feature strongly in this discourse, the overall tenor between public and scientific discourse differs. While public discourse is generally alarmist about the dangers of disinformation in digital communication, scientific discourse is much more skeptical (Altay et al., 2021; Nyhan, 2020; Simon & Camargo, 2021). There is agreement that digital communication environments feature false information, but the degree of their actual reach is strongly contested. The few studies that test the reach of digital information sources suspected of featuring false information indicate only a limited reach among online audiences (Guess et al., 2019) and that it constitutes a very limited share of people’s overall information consumption (Allen et al., 2020). On top of that, a recently disclosed error in the preparation of data sets provided by Facebook for selected research teams suggests that previous work based on these data sets likely overestimates the prevalence of false information on the platform (Alba, 2021). This indicates that disinformation reaches people predominantly through media coverage and public discourse rather than digital media. Correspondingly, some scholars took a critical turn regarding the

¹ We focus explicitly on the perceived threats of digital disinformation in Western democracies. Dangers of disinformation have also been discussed regarding countries that lack strong media institutions, such as Myanmar and the Philippines; countries experiencing long-term, multi-channel foreign influence operations, such as Taiwan; and countries at war, such as Ukraine. In specific circumstances like these, digital disinformation can be expected to carry greater risk and clearly needs to be taken seriously. Our argument does not extend to these cases but focuses on the perceived impact of disinformation in Western democracies as a society-wide phenomenon.

² The terms disinformation, misinformation, and fake news are best understood as distinct concepts. We follow Lecheler and Egelhofer (2022), who define *misinformation* as “incorrect or misleading information” (p. 70) in general; *disinformation* as “incorrect or misleading information that is disseminated deliberately” (p. 71); and *fake news* as both “a type of false information that is the pseudo journalistic imitation of news—it is not only false, but fake” (p. 71) and as a label used “to discredit and delegitimize journalism and news media” (p. 71). These terms figure in discourse warning against unreliable and manipulative information in digital communication environments. We define *disinformation discourse* as public discussion and warnings against disinformation, misinformation, and fake news.

influence of disinformation once the shock of the 2016 U.S. presidential election settled (e.g. Karpf, 2019; Wardle, 2020). Looking at the evidence, indiscriminate and alarmist warnings against disinformation mischaracterize digital communication environments and risk overestimating their effects and associated societal dangers.

By broadly characterizing information encountered in digital communication environments as unreliable and manipulative, the steady stream of alarmist warnings over the last few years against the dangers of digital disinformation may have unintentionally contributed to an overall sense of societal and democratic decline. In fact, several scholars discuss the potential negative effects of alarmist disinformation discourse on perceptions of media and democratic institutions (Benkler et al., 2018; Egelhofer et al., 2020; Jungherr & Schroeder, 2022; Karpf, 2019; Nisbet et al., 2021; Ross et al., 2022; Scheufele & Krause, 2019).

Building on these concerns, we argue that balanced accounts documenting the presence of digital disinformation, while also accounting for empirically established limits of reach and persuasion, offer a promising alternative to indiscriminate and alarmist warnings. By recognizing the presence and dangers of digital disinformation while at the same time contextualizing them with state-of-the-art scientific evidence, we can expect balanced accounts to lower indiscriminate fears and ameliorate negative downstream effects that lower public trust in democratic systems overall.

We test these expectations through a preregistered experiment in the United States. We test the effects of two treatments designed to resemble journalistic contributions to disinformation discourse: one emphasizes the dangers of disinformation indiscriminately and in alarmist terms (T1); the other provides a balanced account of what is known about disinformation and its role in society (T2).

The indiscriminate, alarmist treatment emphasizing the dangers of disinformation (T1) raised the perceived dangers of disinformation among recipients. In contrast, the balanced treatment (T2) lowered the perceived threat level. T1, but not T2, had negative downstream effects, increasing respondent support for heavily restrictive regulation of speech in digital communication environments. With respect to support for democratic practices and democracy overall, neither T1 nor T2 had any significant effects. T1 did, however, lower satisfaction with the current state of democracy. Overall, we see among all respondents a positive correlation between the perceived threat of disinformation to societies and skepticism toward democracy.

Our article offers important evidence to better understand the impact of disinformation by focusing on the role of alarmist discourse. Our findings clearly show that there are negative downstream effects of alarmist warnings against the threat of disinformation. These effects are rather small but still substantial given the experiment context. This is particularly so since we used only one short article as treatment and encountered an exceptionally high preexisting baseline within the population regarding the perception of the problem, as expressed in the control group. These negative effects highlight the importance of disinformation scholars, public communicators, and journalists to reflect the breadth of scientific evidence and not accidentally increase the risks for democracy through unbalanced warnings.

Effects of Alarmist Warnings in Disinformation Discourse

Disinformation discourse tends to be alarmist (Carlson, 2020; Farkas, 2023; Farkas & Schou, 2019; Gutsche, 2018). The effects of alarmist warnings are still ill-understood, but related findings underline their potential importance and unintended consequences.

Various studies show that the term “fake news” is best understood as an umbrella term to characterize information, news reporting, or elite communication as unreliable, intentionally misleading, or false. The use of the term “fake news” tends to be antagonistic, performative, and accusatory; speakers use the term to actively delegitimize an “other”—such as the news media, digital news sources, and political opponents. Crucially, the term serves as a label to challenge journalistic and political legitimacy. Similarly, the continual emphasis on the dangers of disinformation in digital communication environments by journalism and political elites serves to delegitimize the challenge digital media represent to established discursive and political elites (Carlson, 2020; Egelhofer et al., 2020, 2022; Egelhofer & Lecheler, 2019; Farhall et al., 2019; Farkas & Schou, 2018; Jungherr & Schroeder, 2022; Li & Su, 2020; Meeks, 2020; Tong et al., 2020). Therefore, the discourse about disinformation is not solely—or perhaps even primarily—an expression of concern about information quality; rather, it is about the political or institutional performance of one’s own legitimate superiority over other discursively competing actors. The label “fake news” and alarmist warnings against disinformation go beyond single sources or instances of information; they can lead people to distrust epistemic institutions in general (Benkler et al., 2018; Farrell & Schneier, 2018; Karpf, 2019; Scheufele & Krause, 2019).

Examining downstream effects of alarmist warnings in disinformation discourse is still rare, though. Jones-Jang et al. (2021) found in a panel study in the United States that higher levels of perceived exposure to misinformation predicted higher levels of political cynicism at a later point. Nisbet et al. (2021) showed in a cross-sectional survey that among U.S. respondents, the perceived influence of misinformation on others was correlated with lowered satisfaction with the state of democracy. This shows that the perception of widespread misinformation affects how people assess politics beyond the narrow confines of news or information.

Most available evidence examines the effects of disinformation experience and discourse on the credibility of information or sources. Studies have tested the effects of actual exposure to misleading information (Altay et al., 2022; Vaccari & Chadwick, 2020), exposure to general warnings about misleading information (Ternovski et al., 2021; Van Duyn & Collier, 2019), exposure to labels indicating specific informational items to be false (Jahng et al., 2021), and the general perception of misleading information being a problem (Müller & Schulz, 2019; Nisbet et al., 2021; Stubenvoll et al., 2021). Evidence from panel and experimental designs indicates that exposure to these forms of disinformation experience or discourse has a negative effect on various credibility assessments and trust. For example, exposure to warnings about disinformation lowered the degree to which people assessed information to be credible, irrespective of actual factualness (Freeze et al., 2021; Jahng et al., 2021; Ternovski et al., 2021; van der Meer et al., 2023; Van Duyn & Collier, 2019), lowered the credibility of sources (Altay et al., 2022), or contributed to people con-

sidering the news media in general to be less credible (Stubenvoll et al., 2021; Vaccari & Chadwick, 2020; Van Duyn & Collier, 2019). Disinformation experience and discourse thus clearly have a negative impact not only on credibility assessments of information but also on the epistemic institutions that produce information.

We build on these findings and extend their reasoning. While previous studies looked predominantly at the effects of specific warnings about disinformation in specific information items or sources (for exceptions, see Jones-Jang et al. (2021) and Ross et al. (2022)), we examine the effects of indiscriminate and alarmist warnings about the dangers of disinformation in general. We also extend the underlying reasoning by focusing not on the effects on trust in sources or the credibility of information, but instead by examining the downstream effects on people with respect to their support for more restrictive regulation of speech (see also Cheng & Chen, 2020; Skaaning & Krishnarajan, 2021) and for liberal democracy (see also Nisbet et al., 2021; Ross et al., 2022). These are relevant attitude objects that are clearly, if indirectly, linked to public debates about the reliability or dangers of widely used information environments in democracy.

First, though, we focus on the direct effect of alarmist warnings against disinformation on the perception of disinformation as a societal threat:

H1a: We expect indiscriminate warnings against the dangers of disinformation (T1) to raise the perceived threat of disinformation.³

H1b: Conversely, balanced accounts of the threats of disinformation (T2) should lower the perceived threat of disinformation.

For H1a, it is important to note that current high levels of perceived dangers of disinformation might serve as a ceiling for the treatment to further add to these fears. Accordingly, the hypothesized effect may be too small to be observable, depending on the population-wide threat perception.

Liberal democracy depends on the free exchange of competing, contradicting, controversial, and sometimes offensive views. Conflict and competition between views and groups are crucial features of democracy that surface different views and interests, indicate their relative strength, and allow for public negotiation between views and interests (Dahl, 1989), and are central to the perceived capacity of democracies to solve complex challenges (Landemore, 2012). For this to work, however, people living in democracies must accept the legitimacy of views held by others, even if they do not agree, those people's political rights to participate politically, and accept political conflict and competition itself (Sullivan & Transue, 1999). They need to be willing to be challenged in their views by others and accept that their preferred candidates may lose elections (Przeworski, 2019).

These preconditions are challenged when people start to believe supporters of other parties are gullible and easily fall for disinformation (Altay & Acerbi, 2022; Nisbet et al., 2021; Ross et al., 2022). Why should I accept electoral defeat, when

³ The hypothesis has been edited for clarity from its original form in the preregistration, without changing its meaning.

I am convinced that voters were misled and acted against their better interest? Why should I listen to the other side, when I am convinced that their views are the result of skillfull manipulation? Alarmist discourse about the threat of disinformation could thus indirectly weaken democracy by delegitimizing conflict and the views of others by giving people on either political side the pretext of discounting divergent opinions as the result of manipulation instead of legitimiante political differences. This would weaken support for crucial elements of liberal democracy and the legitimacy and bindingness of electoral defeat.

Accordingly, we expect indiscriminate warnings against the dangers of disinformation (T1) to

H2a₁ lower support for democratic principles, that is, unconditional acceptance of election results and (H2a₂) support for pluralistic debate.⁴

For T2, we expect no effects along this dimension (H2b_{1,2}).

In the United States, consistent warnings about the dangers of disinformation have stimulated widespread speculation about election manipulation (Coppins, 2020) and foreign influence (Santariano, 2019). Accordingly, alarmist warnings against disinformation can be expected to impact attitudes about the foundational institutions of liberal democracy itself. We follow insights from the rich literature on attitudes toward democracy and differentiate between two different but related aspects (Claassen & Magalhães, 2021). Exposure to alarmist warnings against disinformation could lead to a lowering of the situational assessment of how well democracy currently functions as a system of governance. Still, general support for democracy as a system of governance might be more robust. Attitudes toward liberal democracy are often symbolically charged, deeply held, and connected with people's sense of self. The effect on these attitudes should be weaker than those on the situational evaluation of the workings of democracy.

We expect indiscriminate warnings against the dangers of disinformation (T1) to

H3a lower satisfaction with the way democracy currently functions in the United States.

H4a lower support for democracy as a system of governance.⁵

However, given that support for democracy as a system of governance has been shown to be comparatively stable in other contexts (Claassen & Magalhães, 2021), we expect that the effect hypothesized in H4a may not be observable.

Furthermore, for T2, we expect no effects along these dimensions (H3b–H4b).

⁴ The hypothesis has been edited for clarity from its original form in the preregistration, without changing its meaning.

⁵ The hypothesis has been edited for clarity from its original form in the preregistration, without changing its meaning.

Free and impartial information infrastructures are an essential element of liberal democracies (Müller, 2021). This function was once the exclusive purview of journalism, but today includes digital platforms that have become important distribution hubs for information (Jungheer & Schroeder, 2021), which makes these platforms objects for regulation. While people living in liberal democracies generally tend to support a free press with only light regulatory or governmental interference, it is not clear whether this also translates into governance preferences toward digital platforms (Martin & Hassan, 2022; Mitchell & Walker, 2021). In fact, attitudes toward the governance of and liability for perceived harms of digital platforms are very likely impacted by the discourse on the dangers of disinformation in digital communication environments.

Alarmist disinformation discourse foregrounds accounts of digital platforms being overwhelmed by malicious actors, having the functioning of their own platforms turned against themselves, or simply being disinterested in recognizing or engaging the problem. Exposure will translate into greater demand for government control of digital information infrastructures, even at the cost of limiting the rights to and protections of free speech (Bazon, 2020). In fact, this heightened perception of unruly communication environments may also increase the demand for greater government control over traditional news media.

H5a We expect indiscriminate warnings against the dangers of disinformation (T1) to heighten support for more restrictive regulation of speech and information in digital communication environments.

H5b For T2, we expect no effects along these dimensions.

Information quality is closely associated with the question of control over communication infrastructures. The rise of digital media has led to an increase in the variety of voices available to audiences and a decrease in the power of gatekeepers to vet information quality (Jungheer et al., 2020). This, in turn, has led to growing awareness of the dangers of unvetted information in political discourse and interest in epistemic institutions, fact checkers or other interventions to determine the factualness of content or sources, as well as support for allowing the labeling of suspicious or problematic content (Bennett & Livingston, 2021; Southwell et al., 2018).

By emphasizing the prevalence of false or actively misleading information in digital communication environments, alarmist disinformation discourse is likely to create public awareness of the need for access to information of high quality in digital communication environments. This could translate into greater demand for information from traditional journalism or politicians.

H6a We expect indiscriminate warnings against the dangers of disinformation (T1) to heighten support for privileging speech and information by representatives of institutions, i.e. politicians and journalists.

H6b For T2, we expect no effects along these dimensions.

Methods

To test these hypotheses, we designed a between-subjects survey experiment with two treatments resembling news articles. The experiment was preregistered at OSF⁶ and approved by the IRBs at the host institutions of both authors. We used a control group ($n=604$) in which respondents were not exposed to any information but were simply surveyed for their attitudes, which we measured as outcome variables. Their responses provide a baseline by which to measure the effects of disinformation discourse.

In treatment group 1 (T1) ($n=596$), we exposed respondents to a treatment mimicking an alarmist journalistic article that emphasizes perceived dangers of disinformation in online communication environments. The treatment presents findings from academic studies by Vosoughi et al. (2018), Mitchell et al. (2020), and Thorson (2016) that are often cited to illustrate the dangers of disinformation and discusses real assessments and reactions by governments. The treatment mirrors actual contributions to disinformation discourse and contains no misleading information or deception of respondents (for a replica of T1, see Online Appendix A.1). In treatment group 2 (T2) ($n=590$), which tests a balanced account of the dangers of disinformation by recognizing its existence while also providing necessary context, we exposed respondents to a treatment mimicking a journalistic article that provides a balanced assessment of the dangers of disinformation in online communication environments. It contains references to the same academic studies as T1 with almost the same wording, but balances them with findings from other oft-cited studies by Allen et al. (2020), Grinberg et al. (2019), and Guess et al. (2019) that speak to the limited reach of online disinformation and its limited effects. While T2's tenor varies from T1 and it provides additional facts beyond those in T1, the treatments are similar with respect to the flow of their arguments, and thus they do not deviate significantly beyond the variations subject to the experiment (for a replica of T2, see Online Appendix A.2).

T1 and T2 allow us to compare the effects of alarmist and balanced warnings. Had we instead chosen genuine journalistic accounts, we would have risked contaminating the effects of these warnings through other features of the chosen artifact. We also could not have compared both treatments directly, as article examples would have differed across other dimensions as well, weakening effect identification.

Both treatments feature no deception. Upon completion of the survey, respondents in both groups were debriefed and told that they had been part of an experiment on the impact of disinformation discourse. They were further told that the debate regarding the effects of disinformation is ongoing.⁷

⁶ Available at: https://osf.io/t8p6k/?view_only=872bd8f34697401c973c721f5872bf5a. All reported analyses and hypotheses follow the preregistration except when explicitly stated.

⁷ We tested both treatments in a preregistered pre-study (https://osf.io/9xcwf/?view_only=6a1f57fab8eb4e99ad061afd7f07cf40) to check whether the intended framings of articles were perceived by participants as intended. The sample size for the pre-study was 150 (around 50 participants per group). Participants were also recruited by *Prolific*. Participants evaluated the framing of the two articles as planned. We asked them how much of a problem the article says made-up news and information are in the United States today (1- No problem at all; 7-Extremely big problem). Participants who were shown the article with indiscriminate warnings against the dangers of disinformation chose, on average, a higher

We recruited 1,800 participants from the survey research company *Prolific*. To estimate the required sample size that would give us sufficient statistical power to test the hypotheses (H1-H6), we ran simulation-based power analyses with outcome variable means and standard deviations identified in the preregistered pre-study. A sample size of 1,800 gives us enough power to identify small effects (see Online Appendix B.1). Data collection ran on 9 May 2022.

U.S.-based persons aged 18 or older could participate. *Prolific* filtered out participants from the pre-study. Since we ran the survey through *Prolific's* European platform, we paid participants the \$ equivalent of £1 for their participation (an hourly rate of £8.57). Following parameters defined in the preregistration, we excluded ten participants who failed a simple attention check at the beginning of the study, resulting in a sample of 1,790.

We assigned participants randomly to one of the three conditions: those in the control group continued directly with questions measuring the outcome variables; those in either treatment group saw the pertinent treatment article. After 30 s, respondents could click to continue with the questionnaire. As a manipulation check, we asked two multiple-choice questions with three answer options and a “I don’t know” option (correct answer for topic of treatment: T1=99.8%, T2=99.7%; correct answer for institution mentioned in treatment: T1=73.5%, T2=77.3%). After answering the questions that measured the dependent variables, participants answered sociodemographic questions. Upon completing the questionnaire, participants received a debriefing and returned to *Prolific*.

As we used an online panel, our sample represents an internet population of which 49.4% was male; the median age was 35 ($M=38.45$, $SD=13.48$); 78.2% reported an annual income of less than US\$100,000; 73.4% identified as white; and 17.8% indicated earning a master’s degree or higher. This means the sample is on average younger, has lower income, and is slightly less white than all Americans (see Online Appendix B.2). These deviations in our sample from the general population are not overly problematic for a survey experiment. Randomization controls show that treatment and control groups resemble each other sufficiently and show no systematic deviations. Randomization worked well; we found no significant differences between participants depending on which of the three conditions they were assigned regarding age ($F(2,1785)=0.206$, $p=.81$),⁸ education ($X^2(2, n=1790)=0.38$, $p<.83$), race ($X^2(20, n=1790)=25.52$, $p<.18$), income ($X^2(24, n=1790)=25.72$, $p<.37$), and gender ($X^2(4, n=1790)=3.24$, $p<.52$).

We measured outcome variables on a 7-point scale (see Table 1) or merged to a 7-point mean index. We used a set of established and novel measures to account for different attitudes related directly or indirectly to disinformation discourse (see Online Appendix C.1).

score ($M=6.72$, $SD=0.54$) than participants shown the article with the balanced assessment ($M=2.19$, $SD=1.07$; Welch- $t(68.74)=26.44$, $p<.001$). Treatments were perceived as intended by us. Furthermore, we tested some of the measurements, as they were specifically created for this study. We used information about mean scores and standard deviations to run the power simulations and estimate the required sample size for the main study.

⁸ There are two missing values for age.

Table 1 Overview of variables used for analysis

Variable	M (SD)	n
H1: Perceived threat of disinformation	5.4 (1.52)	1790
H2 ₁ : Unconditional acceptance of election results	4.28 (1.97)	1790
H2 ₂ : Support for pluralistic debate	3.89 (1.83)	1790
H3: General satisfaction with democracy	3 (1.67)	1790
H4: Support for democracy as a system of governance (6 items, $\alpha=0.73$)	4.93 (1.03)	1788 ^a
H5: Support for more restrictive regulation of speech and information in digital communication environments (3 items, $\alpha=0.86$)	4.5 (1.78)	1790
H6: Support for privileged speech in digital communication environments (2 items, $\alpha=0.93$, Spearman-Brown=0.93)	1.82 (1.49)	1790
Gender (1=male)	50%	1790
Education (1=master's degree of higher)	18%	1790
Political orientation	3.08 (1.81)	1790
Interest in politics	4.73 (1.7)	1790
Control		604
T1		596
T2		590

^aFor two respondents we have missing data for at least one of the measured items used for this index

We follow Mitchell et al. (2019) and measure the perceived threat of disinformation with a single question: “How much of a problem do you think made-up news and information are in the country today?”

For unconditional acceptance of election results and support for pluralistic debate, we constructed two new items that paired a statement of support for one central tenet of democratic theory (i.e. acceptance of electoral defeat and free exchange and debate) with a counterweight indicating that a relevant political other might have been misled in voting or debating (“It is important to accept election results, even if the winning side repeatedly presents arguments that are based on information that has been proven to be false”; and “Democracy works best if there is a free exchange of different political opinions, whether they are based on proven fact or not”). We did so to make support for these statements less of a truism than an assessment of underlying values, even given the threat of disinformation.

We measured general satisfaction with democracy with a single standard item, as in Magalhães (2014) (“I am satisfied with the way democracy works in the US.”). We follow a well-established scale as discussed by Magalhães (2014) in capturing support for democracy as a system of governance with six items (“Having a strong leader who does not have to bother with parliament and elections would be a very good way to govern this country”; “Having experts, not governments, make decisions according to what they think is best for the country would be a very good way to govern this country”; “Having a democratic political system is a very good way to govern this country”; “Democracies are indecisive and have too much squabbling”; “Democracies aren’t good at maintaining order”; and “Democracy may have problems but it’s better than any other form of government”).

To measure support for more restrictive regulation of speech and information, we slightly adjust items previously used by Mitchell and Walker (2021) (“The government should take steps to restrict false information online, even if it limits

people from freely publishing or accessing information”; “Technology companies like Facebook, Twitter or WhatsApp should be held accountable for the behavior of their users, even if this leads them to restrict people from voicing their opinion”; and “Technology companies like Facebook, Twitter or WhatsApp should be held accountable for the behavior of their users, even if this leads them to restrict people from voicing their opinion”). These questions present a tradeoff between stronger control by both government and technology companies and freedom of information, and should make respondents aware of the potential costs associated with an increase in interferences. We added two items covering support for privileging speech and information by representatives of institutions (“Statements by politicians should be exempt from fact checking, labeling, or take downs on digital media”; and “Information provided by news organizations should be exempt from fact checking, labeling, or take downs on digital media”) Both items are original additions but closely follow policy suggestions with respect to privileging selected speech in digital communication environments.

Before treatment exposure, we measured political orientation, education, gender, and political interest. The regression models included these preregistered covariates.⁹

The mix of established and specifically developed items allows us to capture the potential downstream effects of alarmist disinformation discourse more broadly than by relying exclusively on established measures. Still, explicitly connecting specifically developed items to conceptual and theoretical considerations speaks in favor of their fit.

Results

Descriptive statistics of attitudes in the control group provide a first overview of the perception of disinformation as a threat to society among U.S. respondents. There is already a very high threat perception of disinformation among those in the control group, without any prompt by our treatment. When asked to rate the threat of disinformation to U.S. society on a scale from 1 to 7 (1 being low and 7 high), the average was 5.6 (SD=1.41), and the median was 6.

Clearly, the perception of disinformation as a threat is widespread among our respondents. This also indicates that the downstream effects of the treatment providing indiscriminate warnings against disinformation might be somewhat more muted than originally expected. The high levels of threat perception of disinformation introduce a ceiling that exposure to a single treatment might not raise effectively. Consequently, the negative downstream effects of disinformation discourse might already be sufficiently diffused in society through multi-year continuous treatment. Our treatment, therefore, is probably best understood as adding marginally to these effects.

We will first report the correlation between the perception of disinformation as a serious threat to society and the subsequent dependent variables in our control group.

⁹ Data and analysis scripts available at https://osf.io/ycmsw/?view_only=973c62e0d35243379d11e3cfd8543e9d.

The discussion of disinformation is highly politicized, with different political groups accusing each other of slinging disinformation. We find a small correlation between the perceived threat of disinformation and political orientation ($r(602)=-0.18, p<.001$). The more conservative a person is, the lower the perceived threat of disinformation. This result is also supported when we consider party identification. Participants who identify as Republicans ($M=5.20, SD=1.65$) indicate a lower threat perception compared to those who do not ($M=5.71, SD=1.31, t(602)=3.79, p<.001$).

For the correlational analysis, we also added a variable to the questionnaire asking respondents whether politicians who represent people like them are often falsely attacked for spreading misinformation (1- Strongly disagree; 7- Strongly agree). Our data show that the more conservative a person is, the higher the agreement with the statement that politicians representing them are often falsely attacked for spreading disinformation ($r(602)=-0.26, p<.001$). Furthermore, we find that Republicans demonstrate higher agreement with the statement ($M=4.47, SD=1.73$) compared to those who do not identify as Republicans ($M=3.34, SD=1.66, t(602)=6.95, p<.001$).

The perceived threat of disinformation also correlates positively with support for more restrictive regulation of speech and information in digital communication environments ($r(602)=0.37, p<.001$) and negatively with general satisfaction with democracy ($r(602)=-0.18, p<.001$) but not with support for democracy as a system of governance ($r(601)=-0.06, p=.17$). While correlation can only tell us so much about the connection between two variables, these results already indicate that disinformation threat perception relates to negative assessments of the quality of democracy. Now, how do these perceptions shift once we expose respondents to indiscriminate and alarmist warnings against disinformation (T1) and balanced assessments containing information about the dangers of disinformation but also their limited reach and persuasive appeal, as discussed in the scientific literature (T2)?

We tested the hypotheses with OLS regression models. We report results with the Lin (2013) covariate adjustment.¹⁰ Figure 1 shows the results of both treatments.¹¹

First, we checked for both treatments' effect on recipient perceptions of disinformation as a societal problem. According to Hypothesis 1a, the treatment warning indiscriminately of the dangers of disinformation (T1) should increase respondents' sense of threat. In contrast, Hypothesis 1b expects the treatment to present a balanced account of what is known about the limited reach of digital disinformation (T2) to lower respondent threat perception. We find both hypotheses to be supported by the data.

Respondents exposed to indiscriminate warnings of disinformation (T1) expressed a higher problem perception of disinformation for society than those who were not exposed to any treatment (C) ($b=0.17, p=.02, 95\% \text{ CI } [0.03, 0.32]$). In contrast, respondents exposed to the balanced account of digital disinformation and their lim-

¹⁰ This is a deviation from the preregistration. We also report the complete preregistered OLS without the Lin (2013) covariate adjustment, the results without the preregistered covariates, as well as a specification curve analysis. All results are the same, independent of the chosen approach (see Online Appendix D.1–D.3).

¹¹ See Online Appendix D.1 for complete models.

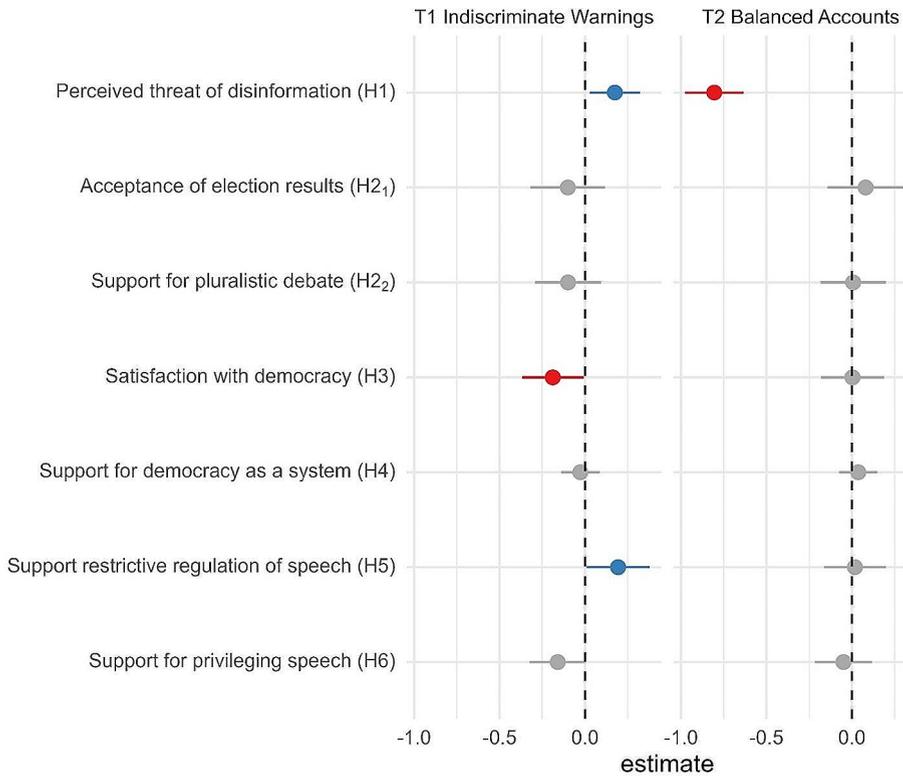


Fig. 1 Estimates with 95%-CI. Left column: T1 indiscriminate warnings; Right column: T2 balanced assessment. A grey line indicates overlap with 0. Color indicates significant positive (blue) or negative effect (red). All identified effects are consistent with our hypotheses

ited reach in society (T2) expressed a significantly lower problem perception than the control group ($b = -0.81, p < .001, 95\% \text{ CI } [-0.98, -0.63]$).

It is notable that the balanced treatment had a much stronger effect in lowering the threat perception than the indiscriminate warning had in raising it. Coupled with the absolute values of the threat assessment in the control group, this shows that the threat perception of disinformation is already high in society. We see that the indiscriminate warning hits a ceiling in raising the problem perception. Conversely, the balanced account lowers the problem perception considerably, although it contains the same information as the indiscriminate warning.

The first set of potential downstream effects of disinformation discourse concerns people’s attitudes toward democracy. We first tested for attitudes toward important principles of liberal democracy: the unconditional acceptance of election results and support for pluralistic debate. Other than expected (H2a_{1,2}), the indiscriminate warning of disinformation (T1) did not impact attitudes toward both the unconditional acceptance of election results ($b = -0.10, p = .37, 95\% \text{ CI } [-0.32, 0.12]$) and support for pluralistic debate ($b = -0.1, p = .31, 95\% \text{ CI } [-0.29, 0.09]$). As expected (H2b_{1,2}),

the balanced warning (T2) also did not show effects (acceptance: $b=0.08$, $p=.48$, 95% CI [-0.14, 0.30]; pluralistic debate: $b=0.01$, $p=.94$, 95% CI [-0.18, 0.20])

Beyond the principle of democracy, disinformation discourse may also impact levels of satisfaction with the current state of democracy. As expected (H3a), the indiscriminate warning of disinformation (T1) had a significant negative effect on satisfaction with the current state of democracy in the United States reported by respondents when compared with respondents who were not shown information about disinformation (C) ($b=-0.19$, $p=.041$, 95% CI [-0.37, -0.01]).¹² As expected (H3b), the balanced discussion (T2) did not have a significant effect ($b=0.003$, $p=.97$, 95% CI [-0.18, 0.19]).

We also tested the effect of disinformation warnings on support for democracy as a system of governance. Here, following the literature on attitudes on democracy, expectations were muted, given the stability of support for the principle of democracy (H4a, H4b). Corresponding to findings from this larger literature, we found that both the indiscriminate warning (T1) ($b=-0.03$, $p=.64$, 95% CI [-0.14, 0.09]) and the balanced account (T2) ($b=0.04$, $p=.52$, 95% CI [-0.08, 0.15]) did not show effects on respondents attitudes toward democracy as a system of governance.

Beyond democracy itself, the perception of disinformation in digital communication environments being a threat to society can be expected to lead to greater acceptance of regulatory interventions. While digital communication environments demand regulation of speech and user behavior, regulatory overreach may prove another danger to democracy, even if somewhat underreflected at present. As expected (H5a), we found the indiscriminate warning against the dangers of disinformation (T1) to increase support for more restrictive governance of digital communication environments ($b=0.19$, $p=.040$, 95% CI [0.01, 0.38]). Also as expected (H5b), the balanced assessment (T2) did not lead to this effect ($b=0.02$, $p=.085$, 95% CI [-0.16, 0.20]). While regulation of these environments clearly is important, public support for restrictive regulation at the cost of political expression is troubling and is potentially detrimental for democracy and democratic discourse.

Finally, we tested the effect of warnings of disinformation on the support among respondents for privileging speech by journalists and politicians in digital communication environments. Against our expectation (H6a), indiscriminate warnings of disinformation (T1) did not show an effect on these assessments ($b=-0.16$, $p=.057$, 95% CI [-0.32, 0.00]). The same was true, as expected, for the balanced warning (T2) ($b=-0.05$, $p=.564$, 95% CI [-0.22, 0.12]).

Discussion and Conclusion

The perception of disinformation as a serious threat to society is quite prevalent among our respondents. This perception, notably, is correlated with dissatisfaction with the current state of democracy in the United States and preferences for restrictive regulation of speech on platforms. Evidence from the experiment further demon-

¹² While significant, the effect was only just above the significance level. Future research should examine the robustness of this finding.

strates that indiscriminate warnings against digital disinformation raise the problem perception of disinformation as a societal threat, lower satisfaction with the current state of U.S. democracy, and increase support for restrictive regulation of digital communication environments. In contrast, balanced explanations of digital disinformation that account for its presence in digital communication environments, provide evidence of its comparatively limited reach, and show its limited persuasive appeals lowered threat perception and did not contribute to negative downstream effects regarding satisfaction with democracy or support for restrictive regulation.

These findings come with some limitations. The reach and contexts of threats of disinformation vary among countries (Humprecht et al., 2020). Our study, of course, concerns only the United States, which is clearly a special case with its high level of partisan polarization, the readiness of prominent political elites to initiate and reinforce disinformation sources and beliefs, and frequent public debate about the dangers of disinformation with across-the-aisle accusations of political others being slingers or at least dupes of disinformation. But even in this context saturated with disinformation warnings, we see alarmist discourse adds to fears about disinformation and enforces downstream effects. These patterns should be even more pronounced in countries with low levels of alarmist disinformation discourse. Examining these variations systematically is not in the scope of the present article, but offers promising opportunities for future research.

Further, our study reported only small effects. We expected this, as participants were exposed only to single informational treatments (for preregistration and power analysis, see Online Appendix B.1). Here, our takeaway is that the treatments showed effects even after this single exposure. As the public discourse on disinformation is a continuous phenomenon, and people are exposed to warnings repeatedly from many sources on many channels, we expect the effects identified here to accumulate over time, not the least because increased threat perception can also have long-term effects, as Jones-Jang et al. (2021) show. Accordingly, it would be promising for future research to test for the cumulative effects of repeated exposure to different forms of disinformation discourse, as well as their duration and long-term effects (Hoes et al., 2022).

It is also interesting to note that the alarmist warning (T1) brought negative downstream effects while the balanced account (T2) showed none. Since T2 lowered people's threat perception, one might have expected T2 also to improve people's satisfaction with the state of democracy and lessen support for restrictive speech regulation. Our findings do not allow us to speak confidently on the reasons for this finding, but one might be that alarmist warnings create a stronger response for increasing protections than a balanced account does for loosening them. In other words, the demand for increasing regulation may react more elastically to threats than the support for loosening regulation reacts to balanced accounts. This is a promising question for further research on attitudes toward regulation and state interference generally.

More broadly, the reported findings inform on effects and inherent risks to the public discussion of disinformation. There is a tradeoff for communicators when discussing the threat of disinformation. Intentional, organized, and coordinated disinformation clearly is a feature of contemporary information environments, including both their digital and their traditional components. This makes it an important

question for journalists, politicians, regulators, and scientists to address publicly. But these warnings can come at a cost.

Our findings show that it matters how the threat of disinformation is communicated. Indiscriminate and alarmist warnings clearly raise public threat perceptions, but also carry negative downstream effects regarding attitudes toward democracy and regulation. The correlative evidence of the connection between high public threat perceptions and negative assessments of democratic satisfaction and support for restrictive regulation of speech online is troubling. At the same time, the evidence regarding balanced accounts of digital disinformation, accounting for their presence as well as the limits of their influence, proves encouraging. It is clearly possible to speak publicly about disinformation without risking raising unfocused fears and stimulating negative downstream effects regarding democratic satisfaction or public willingness to curtail speech in digital communication environments. This shows the importance for public communicators such as journalists, politicians, regulators, and scientists to provide contextualized assessments when discussing the threat of disinformation as a way to provide people with access to the current state of scientific debate and available evidence, or at least transparently communicate existing uncertainties.

Of course, public communicators must not downplay or ignore the dangers of disinformation in digital environments and political discourse at large. Democracies suffer when political elites lose their commitment to factually grounded communication. We see this in established democracies such as the United States. While the literature suggests limited reach and effects of online disinformation (Allen et al., 2020; Grinberg et al., 2019; Guess et al., 2019), such studies have their limitations, as some focus on single platforms or examine the reach of a predefined set of disinformation sources. More importantly, they tend to show average effects on the average citizen. This does not negate the significant role disinformation may play within fringe communities. Furthermore, the 2020 U.S. election, for example, has demonstrated that a sizable segment of the population claims to believe the election was stolen from incumbent candidate Donald Trump (Arceneaux & Truex, 2022). However, it should be noted that disinformation campaigns in the context of the 2020 U.S. election were primarily elite-driven and relied heavily on mass media (Benkler et al., 2020). They do not, therefore, serve as a convincing case for the ills of digital information environments.

Furthermore, the role of disinformation should not be underestimated in countries with weaker democratic and media institutions such as Brazil, Myanmar, and the Philippines. The same applies to countries that are subject to strong multi-channel foreign influence operations, such as Taiwan, or countries at war, such as Ukraine (Rauchfleisch et al., 2023). But while disinformation features in all these cases, it does so to different degrees and under different contextual conditions. The contingencies of the impact of disinformation and its digital components on factors such as the strength of existing democratic and media institutions, the behavior of elites, public trust in institutions and media, and political literacy of populations need to feature strongly in research and public communication.

Still, as our findings show, it is clearly possible to discuss the threat of disinformation in context. By embedding warnings of disinformation in contextual information about the reach of disinformation and the context dependency of effects, commu-

nicators can provide warnings without risking negative downstream effects such as weakening satisfaction with democracy or generating support for restrictive regulatory regimes. This contextually aware disinformation discourse depends on empirical findings that assess the actual reach of disinformation within the general population as well as specific at-risk groups, the differential and potentially heterogeneous effects, and a realistic assessment of effect strengths. Disinformation research needs to put more effort into providing evidence on these questions (Camargo & Simon, 2022).

Public discourse on the impact of digital media on democracy and society has matured of late. The early dominance of accounts of the empowering opportunities of digital media has been replaced by the dominance of those emphasizing dangers to democracy and manipulation. This has provided important impulses to the discourse and strengthened public and regulatory awareness of the importance of critically reflecting digital communication environments' state, effects, and dynamics and improving their governance to support a diverse, strong, and vibrant democratic public arena. But, as this article shows, it matters that we remain balanced in our accounts. By exaggerating the dangers of digital communication environments for democracy in public discourse, we may end up damaging the very thing we wish to protect and improve. The stories we tell about digital media and their role in democracy and society matter. We need to choose well and choose responsibly.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11109-024-09911-3>.

Acknowledgements The authors want to thank Julia Bettecken, Scott Cooper, Pascal Jürgens, Gonzalo Rivero, and Alexander Wuttke for helpful comments on the draft. Andreas Jungherr's work was supported by the VolkswagenStiftung. Adrian Rauchfleisch's work was supported by the Ministry of Science and Technology, Taiwan (R.O.C) (Grant No 110-2628-H-002-008-).

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing Interests The authors declare no potential competing interests with respect to the research, authorship, and/or publication of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alba, D. (2021). Facebook sent flawed data to misinformation researchers. *The New York Times*. <https://www.nytimes.com/live/2020/2020-election-misinformation-distortions/facebook-sent-flawed-data-to-misinformation-researchers>
- Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, 6(14), eay3539. <https://doi.org/10.1126/sciadv.aay3539>
- Altay, S., & Acerbi, A. (2022). Misinformation is a threat because (other) people are gullible. *PsyArXiv*. <https://doi.org/10.31234/osf.io/n4qqrj>
- Altay, S., Berriche, M., & Acerbi, A. (2021). Misinformation on misinformation: Conceptual and methodological challenges. *PsyArXiv*. <https://doi.org/10.31234/osf.io/edqc8>
- Altay, S., Hacquin, A. S., & Mercier, H. (2022). Why do so few people share fake news? It hurts their reputation. *New Media & Society*, 24(6), 1303–1324. <https://doi.org/10.1177/1461444820969893>
- Arceneaux, K., & Truex, R. (2022). Donald Trump and the lie. *Perspectives on Politics*, 1–17. <https://doi.org/10.1017/S1537592722000901>
- Bazon, E. (2020). The first amendment in the age of disinformation. *The New York Times*. <https://www.nytimes.com/2020/10/13/magazine/free-speech.html>
- Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.
- Benkler, Y., Tilton, C., Etling, B., Roberts, H., Clark, J., Faris, R., Kaiser, J., & Schmitt, C. (2020). *Mail-In Voter Fraud: Anatomy of a Disinformation Campaign* Berkman Center Research Publication No. 2020-6.
- Bennett, W. L., & Livingston, S. (Eds.). (2021). *The disinformation age: Politics, technology, and disruptive communication in the United States*. Cambridge University Press.
- BVerfG (2021). *Beschluss des Ersten Senats vom 20. Juli 2021–1 BvR 2756/20, 2775/20 und 2777/20 – Staatsvertrag Rundfunkfinanzierung*.
- Camargo, C. Q., & Simon, F. M. (2022). Mis- and disinformation studies are too big to fail: Six suggestions for the field's future. *Harvard Kennedy School Misinformation Review*, 3(5), 1–9. <https://doi.org/10.37016/mr-2020-106>
- Carlson, M. (2020). Fake news as an informational moral panic: The symbolic deviancy of social media during the 2016 US presidential election. *Information Communication & Society*, 23(3), 374–388. <https://doi.org/10.1080/1369118X.2018.1505934>
- Cheng, Y., & Chen, Z. F. (2020). The influence of presumed fake news influence: Examining public support for corporate corrective response, media literacy interventions, and governmental regulation. *Mass Communication and Society*, 23(5), 705–729. <https://doi.org/10.1080/15205436.2020.1750656>
- Claassen, C., & Magalhães, P. C. (2021). Effective government and evaluations of democracy. *Comparative Political Studies*, 55(5), 860–894. <https://doi.org/10.1177/001041402111036042>
- Coppins, M. (2020). The billion-dollar disinformation campaign to reelect the president. *The Atlantic*. <https://www.theatlantic.com/magazine/archive/2020/03/the-2020-disinformation-war/605530/>
- Dahl, R. A. (1989). *Democracy and its critics*. Yale University Press.
- Egelhofer, J. L., & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: A framework and research agenda. *Annals of the International Communication Association*, 43(2), 97–116. <https://doi.org/10.1080/23808985.2019.1602782>
- Egelhofer, J. L., Aaldering, L., Eberl, J. M., Galyga, S., & Lecheler, S. (2020). From novelty to normalization? How journalists use the term fake news in their reporting. *Journalism Studies*, 21(10), 1323–1343. <https://doi.org/10.1080/1461670X.2020.1745667>
- Egelhofer, J. L., Boyer, M., Lecheler, S., & Aaldering, L. (2022). Populist attitudes and politicians' disinformation accusations: Effects on perceptions of media and politicians. *Journal of Communication*, 72(6), 619–632. <https://doi.org/10.1093/joc/jqac031>
- Farhall, K., Carson, A., Wright, S., Gibbons, A., & Lukamto, W. (2019). Political elites' use of fake news discourse across communications platforms. *International Journal of Communication*, 13, 4353–4375.
- Farkas, J. (2023). Fake news in Metajournalistic Discourse. *Journalism Studies*, 24(4), 423–441. <https://doi.org/10.1080/1461670X.2023.2167106>
- Farkas, J., & Schou, J. (2018). Fake news as a floating signifier: Hegemony, antagonism and the politics of falsehood. *Javnost – The Public*, 25(3), 298–314. <https://doi.org/10.1080/13183222.2018.1463047>

- Farkas, J., & Schou, J. (2019). *Post-truth, fake news and democracy: Mapping the politics of falsehood*. Routledge.
- Farrell, H., & Schneider, B. (2018). *Common-knowledge attacks on democracy*. The Berkman Klein Center for Internet & Society.
- Freeze, M., Baumgartner, M., Bruno, P., Gunderson, J. R., Olin, J., Ross, M. Q., & Szafran, J. (2021). Fake claims of fake news: Political misinformation, warnings, and the tainted truth effect. *Political Behavior*, 43(4), 1433–1465. <https://doi.org/10.1007/s11109-020-09597-3>
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on twitter during the 2016 U.S. Presidential election. *Science*, 363(6425), 374–378. <https://doi.org/10.1126/science.aau2706>
- Guess, A., Nagler, J., & Tucker, J. A. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), eaau4586. <https://doi.org/10.1126/sciadv.aau4586>
- Gutsche, R. E. (2018). News boundaries of fakiness and the challenged authority of the press. In R. E. Gutsche (Ed.), *The Trump presidency, journalism, and democracy* (pp. 39–58). Routledge.
- High Representative of the Union for Foreign Affairs and Security Policy. (2018). *Action plan against disinformation*. European Commission.
- Hoes, E., von Hohenberg, B. C., Gessler, T., Wojcieszak, M., & Qian, S. (2022). The cure worse than the disease? How the media's attention to misinformation decreases trust. *PsyArXiv*. <https://doi.org/10.31234/osf.io/4m92p>
- 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. <https://doi.org/10.1177/1940161219900126>
- Jahng, M. R., Stoycheff, E., & Rochadiat, A. (2021). They said it's fake: Effects of discounting cues in online comments on information quality judgments and information authentication. *Mass Communication and Society*, 24(4), 527–552. <https://doi.org/10.1080/15205436.2020.1870143>
- Jones-Jang, S. M., Kim, D. H., & Kenski, K. (2021). Perceptions of mis- or disinformation exposure predict political cynicism: Evidence from a two-wave survey during the 2018 US midterm elections. *New Media & Society*, 23(10), 3105–3125. <https://doi.org/10.1177/1461444820943878>
- Jungherr, A., & Schroeder, R. (2021). Disinformation and the structural transformations of the public arena: Addressing the actual challenges to democracy. *Social Media + Society*, 7(1), 1–13. <https://doi.org/10.1177/2056305121988928>
- Jungherr, A., & Schroeder, R. (2022). *Digital transformations of the public arena*. Cambridge University Press.
- Jungherr, A., Rivero, G., & Gayo-Avello, D. (2020). *Retooling politics: How digital media are shaping democracy*. Cambridge University Press.
- Kanno-Young, Z., & Kang, C. (2021). They're killing people: Biden denounces social media for virus disinformation. *The New York Times*. <https://www.nytimes.com/2021/07/16/us/politics/biden-facebook-social-media-covid.html>
- Karpf, D. (2019). On digital disinformation and democratic myths. *MediaWell*. <https://mediawell.ssrc.org/expert-reflections/on-digital-disinformation-and-democratic-myths/>
- Landemore, H. (2012). *Democratic reason: Politics, collective intelligence, and the rule of the many*. Princeton University Press.
- Lecheler, S., & Egelhofer, J. L. (2022). Disinformation, misinformation, and fake news: Understanding the supply side. In J. Strömbäck, K. Åsa Wikforss, T. Glüer, Lindholm, & H. Oscarsson (Eds.), *Knowledge resistance in high-choice information environments* (pp. 69–87). Routledge.
- Li, J., & Su, M. H. (2020). Real talk about fake news: Identity language and disconnected networks of the US public's fake news discourse on Twitter. *Social Media + Society*, 6(2), 1–14. <https://doi.org/10.1177/2056305120916841>
- Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *Annals of Applied Statistics*, 7(1), 295–318. <https://doi.org/10.1214/12-AOAS583>
- Magalhães, P. C. (2014). Government effectiveness and support for democracy. *European Journal of Political Research*, 53(1), 77–97. <https://doi.org/10.1111/1475-6765.12024>
- Martin, J. D., & Hassan, F. (2022). Testing classical predictors of public willingness to censor on the desire to block fake news online. *Convergence*, 28(3), 867–887. <https://doi.org/10.1177/13548565211012552>
- Meeks, L. (2020). Defining the enemy: How Donald Trump frames the news media. *Journalism & Mass Communication Quarterly*, 97(1), 211–234. <https://doi.org/10.1177/1077699019857676>

- Mitchell, A., & Walker, M. (2021). *More americans now say government should take steps to restrict false information online than in 2018*. Pew Research Center.
- Mitchell, A., Gottfried, J., Stocking, G., Walker, M., & Fedeli, S. (2019). *Many americans say made-up news is a critical problem that needs to be fixed*. Pew Research Center.
- Mitchell, A., Oliphant, J. B., & Shearer, E. (2020). *About seven-in-ten U.S. adults say they need to take breaks from COVID-19 news*. Pew Research Center.
- Müller, J. W. (2021). *Democracy rules*. Allen Lane.
- Müller, P., & Schulz, A. (2019). Facebook or Fakebook? How users' perceptions of fake news are related to their evaluation and verification of news on Facebook. *Studies in Communication and Media*, 8(4), 547–559. <https://doi.org/10.5771/2192-4007-2019-4-547>
- Nisbet, E. C., Mortenson, C., & Li, Q. (2021). The presumed influence of election misinformation on others reduces our own satisfaction with democracy. *Harvard Kennedy School Misinformation Review*, 1, 1–16. <https://doi.org/10.37016/mr-2020-59>
- Nyhan, B. (2020). Facts and myths about misperceptions. *Journal of Economic Perspectives*, 34(3), 220–236. <https://doi.org/10.1257/jep.34.3.220>
- Przeworski, A. (2019). *Crises of democracy*. Cambridge University Press.
- Rauchfleisch, A., Tseng, T. H., Kao, J. J., & Liu, Y. T. (2023). Taiwan's Public Discourse about Disinformation: The role of Journalism, Academia, and politics. *Journalism Practice*, 17(10), 2197–2217. <https://doi.org/10.1080/17512786.2022.2110928>
- Ross, A. R. N., Vaccari, C., & Chadwick, A. (2022). Russian meddling in U.S. elections: How news of disinformation's impact can affect trust in electoral outcomes and satisfaction with democracy. *Mass Communication and Society*, 25(6), 786–811. <https://doi.org/10.1080/15205436.2022.2119871>
- Santariano, A. (2019). Facebook identifies Russia-linked misinformation campaign. *The New York Times*. <https://www.nytimes.com/2019/01/17/business/facebook-misinformation-russia.html>
- Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, 116(16), 7662–7669. <https://doi.org/10.1073/pnas.1805871115>
- Simon, F. M., & Camargo, C. Q. (2021). Autopsy of a metaphor: The origins, use and blind spots of the infodemic. *New Media & Society*, 1–22. <https://doi.org/10.1177/14614448211031908>
- Skaaning, S. E., & Krishnarajan, S. (2021). Who cares about free speech? Findings from a global survey of support for free speech. *Justitia*. https://futurefreespeech.com/wp-content/uploads/2021/06/Report_Who-cares-about-free-speech_21052021.pdf
- Southwell, B., Thorson, E. A., & Sheble, L. (Eds.). (2018). *Misinformation and mass audiences*. University of Texas Press.
- Stubenvoll, M., Heiss, R., & Matthes, J. (2021). Media trust under threat: Antecedents and consequences of misinformation perceptions on social media. *International Journal of Communication*, 15, 2765–2786.
- Sullivan, J. L., & Transue, J. E. (1999). The psychological underpinnings of democracy: A selective review of research on political tolerance, interpersonal trust, and social capital. *Annual Review of Psychology*, 50, 625–650. <https://doi.org/10.1146/annurev.psych.50.1.625>
- Ternovski, J., Kalla, J., & Aronow, P. M. (2021). Deepfake warnings for political videos increase disbelief but do not improve discernment: Evidence from two experiments. *OSF Preprints*. <https://doi.org/10.31219/osf.io/dta97>
- Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, 33(3), 460–480. <https://doi.org/10.1080/10584609.2015.1102187>
- Tong, C., Gill, H., Li, J., Valenzuela, S., & Rojas, H. (2020). Fake news is anything they say! – conceptualization and weaponization of fake news among the American public. *Mass Communication and Society*, 23(5), 755–778. <https://doi.org/10.1080/15205436.2020.1789661>
- 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), 1–13. <https://doi.org/10.1177/2056305120903408>
- van der Meer, T. G. L. A., Hameleers, M., & Ohme, J. (2023). Can fighting Misinformation have a negative spillover effect? How warnings for the threat of Misinformation can decrease General News credibility. *Journalism Studies*, 24(6), 803–823. <https://doi.org/10.1080/1461670X.2023.2187652>
- Van Duyn, E., & Collier, J. (2019). Priming and fake news: The effects of elite discourse on evaluations of news media. *Mass Communication and Society*, 22(1), 29–48. <https://doi.org/10.1080/15205436.2018.1511807>
- 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>

Wardle, C. (2020). The media has overcorrected on foreign influence. *Lawfare*. <https://www.lawfareblog.com/media-has-overcorrected-foreign-influence>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.