*Human Communication Research*, 2024, **50**, 404–418
https://doi.org/10.1093/hcr/hqae006
Advance access publication 7 May 2024
**Original Research**

# The polarizing content warning: how the media can reduce affective polarization

**Emily Kubin[1,2], Christian von Sikorski [1,*]**

[1]Department of Psychology, University of Kaiserslautern-Landau (RPTU), Landau, Germany
[2]Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, USA
*Corresponding author: Christian von Sikorski. Email: christian.vonsikorski@rptu.de

## Abstract

Past research suggests that journalists can (unintentionally) exacerbate affective polarization when reporting on growing levels of polarization in society. However, is there a way for journalists to report on the realities of growing political polarization without dividing people further? In our research with five pre-registered experimental studies ($N = 3{,}414$), we develop the *polarizing content warning* which, based on inoculation theory, warns readers that scientific research suggests reading news content about political polarization may drive further affective polarization. Results indicate that the polarizing content warning can be used both with online news articles and on social media sites, and is able to indirectly reduce affective polarization of readers. Additionally, the polarizing content warning is beneficial both when presented alongside news content and beforehand, and reduces readers' perceptions of societal polarization, in turn reducing affective polarization. This warning allows journalists to report on societal polarization without further dividing people.

**Keywords:** affective polarization, perceived ideological polarization, inoculation theory, media, social media

In the United States, political polarization is growing. People are more polarized in their political stances (i.e., ideological polarization; Pew Research Center, 2017) and dislike opponents more (i.e., affective polarization; Finkel et al., 2020). While some level of polarization can be healthy for society, e.g., by increasing political participation (Wagner, 2021) and disrupting the status quo (e.g., Kreiss & McGregor, 2023; Stavrakakis, 2018), polarization can also coincide with markers of troubled democracies, such as growing distrust of the government (Hosking, 2019) and support for political violence (Kalmoe & Mason, 2022). Some scholars suggest certain forms of polarization are especially problematic. For example, theoretical arguments suggest that affective polarization is especially troubling for society (Overgaard et al., 2022), suggesting finding ways to combat growing affective divisions is crucial.

Our political beliefs are not just derived from our day-to-day interactions but also by the media (Garrett et al., 2014; Zoizner et al., 2021). Past research suggests media can exacerbate affective polarization (Kim & Zhou, 2020), with media content (especially pro-attitudinal and partisan media) driving affective polarization (for review, see Kubin & von Sikorski, 2021). One type of journalistic reporting that is especially likely to elicit greater affective polarization is media coverage on societal polarization trends—content that is frequently emphasized by journalists (Fiorina et al., 2005). In a first study, Levendusky and Malhotra (2016) find that when people are exposed to news media about increasing ideological polarization (e.g., suggesting that Americans are divided on key political issues like the economy and gay marriage), they become more affectively polarized (i.e., dislike opponents more). However, the authors point to one strategy that can combat these ill effects—emphasizing where citizens agree. Yet, reporting on the reality of growing divides within society is the job of journalists who are supposed to report the unbiased truth. Thus, can journalists report on political divides without further affectively polarizing citizens?

Levendusky and Malhotra's (2016) research framework is seminal to our understanding of how media may (unintentionally) exacerbate the very problem journalists are reporting on. But why does news reporting on polarization make people more affectively polarized? Based on previous research, we posit that media reports about societal polarization increase people's *perceptions* of how ideologically polarized society is. This notion is supported by established connections between media and increased perceived societal polarization (Yang et al., 2016). Additionally, other work focusing specifically on media reports about polarization shows such content shapes at least some polarization attitudes (Levendusky & Malhotra, 2016).

Further, believing society is more ideologically polarized increases affective polarization and animosity toward opponents—a relationship found in many previous studies (e.g., Druckman et al., 2022; Enders & Armaly, 2019; Moore-Berg et al., 2020; Voelkel et al., 2023). These findings are also in line with other work suggesting our political perceptions drive intergroup conflict (Ahler & Sood, 2018; Lees & Cikara, 2020; Ruggeri et al., 2021), and can play an important role in political decisions (e.g., Spiral of Silence Theory; Noelle-Neumann, 1974; see Matthes et al., 2018). Here we propose that news media reports about societal polarization drive affective polarization by increasing readers' perceptions of society being polarized.

The core aim of this research is to develop a warning system that mitigates against the polarizing effects of news media reports that discuss political polarization. While Levendusky and Malhotra's (2016) framework provides a meaningful first step in finding ways for the media to reduce

**Received:** November 21, 2022. **Revised:** March 23, 2024. **Accepted:** March 30, 2024

affective polarization (i.e., through highlighting areas of agreement between opponents), this approach may not be feasible for journalists who have a duty to report on the realities of growing polarization in American society (Pew Research Center, 2017). Due to these realities, and the relevance of perceptions of polarization for elections, inter-party relationships, and political preferences, we extend the valuable work by Levendusky and Malhotra (2016), to establish a new way for journalists to report on societal polarization without exacerbating affective polarization. We test a *polarizing content warning* approach which is presented with online news content about polarization trends in the United States, as a strategy to indirectly mitigate affective polarization.

In the present research, we conducted five studies extending Levendusky and Malhotra's (2016) framework. We explore whether reporting on political polarization predicts perceptions of polarization in society, subsequently driving affective polarization. Further, we assess a novel *polarizing content warning* tool. This warning extends inoculation theory beyond applications within the misinformation literature (Compton, 2013; McGuire, 1964), applying the theory into the field of polarization research by inoculating people against the potentially affectively polarizing effects of certain types of media content. Thus, the current research extends the range of existing theory while simultaneously exploring a novel and theoretically relevant dependent variable (affective polarization)—two key theoretical pursuits in communication research (e.g., DeAndrea & Holbert, 2017). We find media reports about political polarization make people think societal levels of polarization are greater, which increases affective polarization. We also find consistent support for the benefits of the polarizing content warning (both in news articles and social media) for reducing perceived societal ideological polarization and subsequent affective polarization.

## Media strategies to reduce affective polarization

Previous research suggests media exacerbates partisan division and disdain (Kim & Zhou, 2020). Though evidence for this is mixed, as others have found no connection between media exposure and affective polarization (Wojcieszak et al., 2021). Yet, one systematic review suggests there is a shortage of studies testing whether media can reduce affective polarization (Kubin & von Sikorski, 2021), and understanding the relationship between affective polarization and media is highly complex (Kubin & von Sikorski, 2023). Initial research suggests media interventions can have pro-social benefits, such as reducing the dehumanization of migrants by correcting misperceptions (Moore-Berg et al., 2022), reducing negative perceptions of Muslims by portraying them in more positive terms (Saleem et al., 2015), and that media interventions can even reduce affective polarization (Voelkel et al., 2023; Zoizner et al., 2021).

Levendusky and Malhotra (2016) address another way the media can reduce affective polarization, by emphasizing the places where citizens agree (moderation framing) rather than highlighting disagreement (polarization framing). While this strategy is promising, today (especially in the United States), people are becoming more polarized (e.g., Pew Research Center, 2017), making media reports about growing divisions imperative. While some have questioned the extent to which average Americans are in fact ideologically polarized (e.g., Fiorina & Abrams, 2008), there is agreement that polarization is at least occurring among the political elite (e.g., politicians; Fiorina & Abrams, 2008; Robison & Mullinix, 2016).

Given that the existence of polarization among the political elite is well established in the literature, it makes sense that the media needs to report on polarization (at least to some degree). In line with this, previous research suggests polarized politicians are especially likely to receive more media coverage (e.g., Wagner & Gruszczynski, 2018) and that this coverage matters for mass polarization. For example, when people see media coverage of political conflict (between elites), they exhibit greater intergroup animus (Han & Federico, 2018). Further, many scholars argue that people perceive themselves as being divided (e.g., Fiorina, 2016; Lees & Cikara, 2021), a perception that is key for how people understand and engage with the political ecosystem (e.g., driving affective polarization, Druckman et al., 2022; Enders & Armaly, 2019; Moore-Berg et al., 2020), and a perception that can be exacerbated by the media (e.g., Yang et al., 2016).

We suggest perceived societal ideological polarization is a key mechanism for understanding why media portrayals of polarization drive affective polarization—a mediational relationship supported by the literature. For example, Levendusky and Malhotra (2016) find that the moderation (as compared to polarization) article reduces perceptions of how ideologically polarized society is. Further, previous research suggests that perceived polarization frequently exacerbates affective polarization (e.g., Druckman et al., 2022; Enders & Armaly, 2019). Based on these findings, we suggest a novel theoretical framework for understanding how news media content can (indirectly) drive affective polarization. Specifically, we posit:

> **H1:** *News about political moderation (vs. polarization) will reduce affective polarization via the mediational effect of reduced perceived societal ideological polarization.*

## Developing the polarizing content warning

Inoculation theory (McGuire, 1964) posits we can inoculate people against persuasive ideas by warning them of the coming attack (of ideas) and providing tools to combat them (Compton, 2013). Inoculation theory involves two key mechanisms: 1) a (fore)warning of a threat and 2) counterarguing against the threat (Compton et al., 2021a). This theory has been applied and tested in a variety of domains including politics, health, education, and commercial contexts (e.g., van der Linden et al., 2020). Additionally, meta-analytical findings suggest inoculating news consumers is an effective strategy for resisting persuasive attempts (see Banas & Rains, 2010), especially related to combating misinformation beliefs (Basol et al., 2020; Lewandowsky & van der Linden, 2021; van der Linden, 2020).

The threat of news content discussing polarization driving perceived societal polarization and affective polarization is two-fold. First, it is threatening to perceive societal polarization as more extreme than it may be because such perceptions are known to drive further division and affective polarization—even when such perceptions are inaccurate (e.g.,

Druckman et al., 2022; Lees & Cikara, 2020). Second, it is threatening because increased affective polarization is theorized to harm the health of democracy in fundamental ways by reducing cross-cutting conversations, increasing stereotyping, and exacerbating discord (Overgaard et al., 2022). Here, we argue that inoculation theory can combat these effects by warning about the threats (i.e., the polarizing effects of news content) and counterarguing against these threats (i.e., by arguing against readers becoming more polarized after reading the news content). By both warning about the threat and counterarguing against it, we present a theory driven (and practically relevant) warning for indirectly combating affective polarization. But how should such a warning be presented to news consumers? Previous work suggests warnings presented alongside content are beneficial as the processing of the media content has not yet been completed (Moravec et al., 2020) and that such strategies may even be more effective than pre-bunking techniques (Brashier et al., 2020).

In the current research, we apply inoculation theory to polarization research, to assess whether inoculation principles can warn and counterargue against the potentially polarizing effects of media content. Here we explore novel outcomes, examining how warning about the threat of news content exacerbating polarization, and counterarguing against this threat (i.e., counterargue against readers becoming more polarized), can reduce perceived polarization and subsequently reduce affective polarization. We assess the effectiveness of inoculation warnings presented alongside news content (i.e., with-text) and beforehand (pre-bunking), to reduce beliefs about how polarized society is.

But what kinds of information should be included in warnings? Past research suggests scientific research can shift beliefs about politicized topics (e.g., climate change; van der Linden et al., 2015), and arguments from experts are persuasive (Clark & Evans, 2014). This is supported by persuasion literature emphasizing the importance of source credibility in attitude change (Heesacker et al., 1983; Pornpitakpan, 2004), suggesting persuasive warnings include expert sources (e.g., scientific research). The importance of credible sources is also evident in inoculation theory, which suggests source effects are key for successful inoculation messages (Compton & Pfau, 2005). This suggests that different sources (e.g., online news media vs. news on social media), which are known to differ in perceived source credibility (Besalú & Pont-Sorribes, 2021), may influence the effectiveness of inoculation. In the current research, we develop a warning system that includes expert sources (i.e., findings from scientific research) and considers source effects by assessing the warning with both online news articles and news on social media.

Additionally, past work suggests inoculating the public against certain ideas is possible (e.g., Basol et al., 2020; Compton et al., 2016). But can inoculation also reduce beliefs about how polarized society is? We develop the polarizing content warning which inoculates people against the polarizing effects of media content reporting on societal political polarization. The warning emphasizes expertise and source credibility through teaching readers about real scientific research on the polarizing effects of news content about polarization (e.g., Levendusky & Malhotra, 2016). Teaching people how news content can drive affective polarization, is in line with inoculation research. For example, techniques, such as the *Bad News Game* (Roozenbeek & van der Linden, 2019), suggest that teaching how misinformation (or in our case, polarization) functions is key for reducing the persuasion of that misinformation (or in our case, polarizing content). We test whether it is effective both when presented beforehand (pre-bunking) and alongside media content (i.e., with-text).

We suggest, based on inoculation theory, the polarizing content warning counterargues the threat of news content exacerbating polarization by arguing against readers becoming more polarized after reading the news content. This counterargument may reduce people's propensity to become more polarized (i.e., by reducing the likelihood they will perceive high levels of ideological divisions in society, in turn reducing the likelihood of becoming more affectively polarized (Druckman et al., 2022)). We posit:

> **H2**: *News about polarization presented with a polarizing content warning (vs. no warning) will reduce affective polarization via the mediational effect of reduced perceived societal ideological polarization.*

To test the generalizability of the polarizing content warning in journalistic communication, we ask the following:

> ***Research Question (RQ)*: *Is the polarizing content warning effective both when presented with news content about political polarization and when presented beforehand?***

## Current research

The current research consists of five studies that both theoretically and practically expand Levendusky and Malhotra's (2016) framework and inoculation theory. Study 1 assesses H1, H2, and the RQ by examining whether news reporting on polarization increases affective polarization via increasing perceptions of ideological polarization. Study 2 assesses H1, H2, and RQ with variations in how the polarizing content warning is presented. Studies 3–5 assess H2 and RQ by extending results in the social media context (i.e., news shared on social media platforms). Study 4 tests whether the source of the warning influences effects and Study 5 develops warning systems comparable to those already used by Facebook and Twitter for other purposes. This project received ethics approval from the University of Kaiserslautern-Landau (RPTU) (ID: LEK-399_r). Data, materials, and pre-registrations for all studies are available at: https://osf.io/na3dy/?view_only=c970731583d6439baf4ade59bd033dda.

### Study 1

Study 1 conceptually replicated Levendusky and Malhotra's (2016) results, by testing whether an article highlighting moderation in society (vs. growing polarization) would reduce affective polarization. Based on previous research, we propose a mediation model (Figure 1) where the moderation (vs polarization) article reduces perceptions of societal ideological polarization, thereby reducing affective polarization, testing H1.

Study 1 also takes a first step at extending inoculation theory by assessing whether the polarizing content warning affects previously untested dependent variables (perceived ideological polarization and affective polarization)—key theoretical advancements outlined by DeAndrea and Holbert (2017). We test whether the *polarizing content warning* combats the polarizing effects of online news articles, by reducing

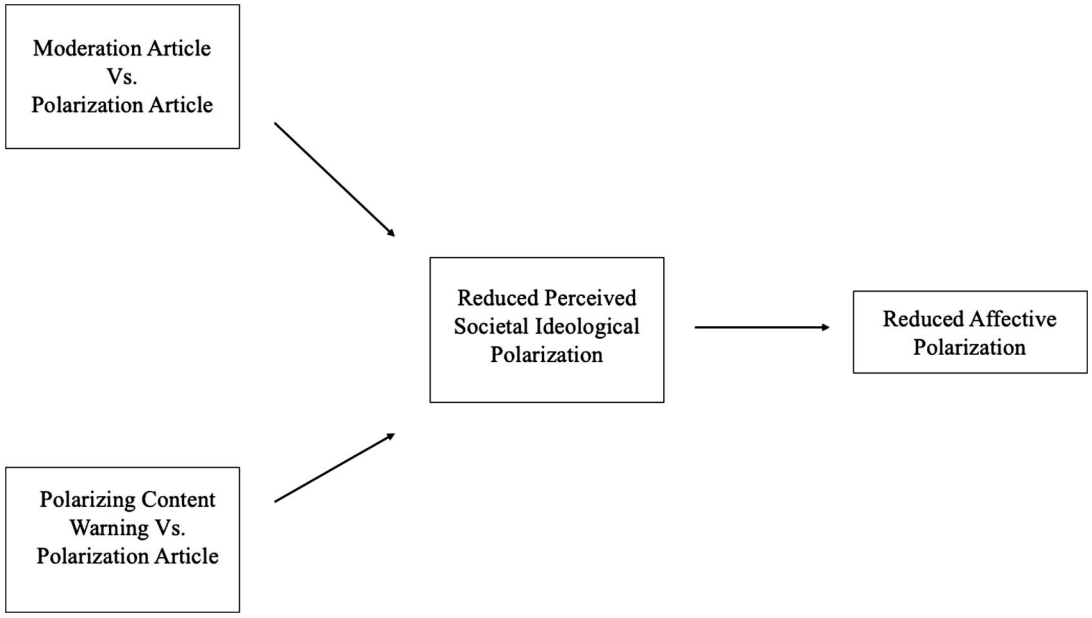


**Figure 1.** Theoretical framework.

perceived societal ideological polarization, in turn reducing affective polarization (testing H2) (See Figure 1). Study 1 also answers the RQ in part, by testing the presentation of the warning *with* news content.

### Sample

American participants ($N = 502$) were recruited from MTurk. As pre-registered, individuals who failed attention or manipulation checks or identified themselves as "true independents"[1,2] were removed from analyses. We removed true independents from all studies as our affective polarization measure focused on warmth toward Democrats and Republicans. We only included people who identified with one of these parties so we could determine their political in-group and out-group. This practice is standard in polarization research (e.g., Iyengar et al., 2012; Levendusky & Malhotra, 2016). After removing these participants, 382 participant responses were included in analyses (Mean Age (SD) = 41.36(12.57), 53.40% Male, 59.69% liberal).

### Procedure

Participants read a news article from USA Today. In the polarization condition, participants saw an article that discussed growing political polarization in the United States. In the moderation condition, participants saw an article that discussed areas where Democrats and Republicans agree. In the with-text polarizing content warning condition, participants saw the identical article as those in the polarization condition; however, there was also a prominent warning label. This label stated that scientific research suggests articles about polarization shift beliefs about how polarized society is which in turn exacerbates affective polarization. See page 2 of Supplementary Materials for stimuli. Participants next responded to the following measures:

#### *Perceived societal ideological polarization*

With this 3-item Likert measure, participants reported the extent to which they feel Americans, "disagree on many policies," "have divergent opinions on political issues," and "are not united in terms of their political views," using a 7-point scale from (1) *strongly disagree* to (7) *strongly agree* ($\alpha = .94$). Responses were averaged together to create a measure of perceived societal ideological polarization, with higher scores indicating greater perceptions of society being ideologically polarized.

#### *Participant affective polarization*

Participants used a 100-point feeling thermometer from 0 (very cold) to 100 (very warm) to rate both Democrats and Republicans—a measure commonly used in political research (e.g., Iyengar, et al., 2012). Based on the party affiliation of participants, we determined whether Democrats or Republicans were part of participants' in-group or out-group. This allowed us to develop a measure for participants warmth ratings toward their political allies and opponents. However, given that growing polarization is primarily driven by outgroup hate (Finkel et al., 2020), we focus on affective polarization warmth ratings toward political opponents. In additional analyses, we also assessed affective polarization as the difference between ones' in-party and out-party warmth ratings. Results across all studies remained consistent using either form of the affective polarization measure. See Supplementary Materials for further information.

### Results

Overall, in both this study, and all other studies, the total effect of mediational analyses was not significant. See Supplementary Materials for further details. However, in line with our hypotheses, and calls from other scholars (e.g., Zhao et al., 2010), it is meaningful to examine indirect effects. All following analyses examine the indirect effects of perceived ideological polarization. Supporting H1, PROCESS mediational analysis[3] (Hayes & Rockwood, 2020) revealed the moderation (1) vs. polarization (0; reference) condition, reduced perceived ideological polarization which reduced affective polarization, indirect effect = 0.44, $SE = 0.08$, 95% CI: 0.28–0.60. The moderation condition directly affected affective polarization ($B = -0.45$, SE = 0.14, 95% CI: −0.73,

−0.17). While H1 focused on the mediational effect, it is noteworthy to see the direct effect running in the opposite direction of the indirect effect (in the mediation model only). This could suggest there is an unstudied second mediator predicting affective polarization (O'Rourke & Mackinnon, 2018). See Figure 2 and Table 1.

The with-text polarizing content warning (1) vs. polarization (0) article reduced perceptions of societal ideological polarization which reduced affective polarization, indirect effect = 0.07, SE = 0.03, 95% CI: 0.02–0.13, supporting H2 and answering RQ1.The direct effect was not significant ($B = -0.11$, SE = 0.12, 95% CI: −0.34, 0.13). Importantly, an indirect effect can also be significant when one of the paths in a mediation model is insignificant or trending toward significance (Igartua & Hayes, 2021). See Figure 2 and Table 1 for path coefficients and model details and Supplementary Table S1 for mean differences by condition.

### Discussion

Results supported H1 and indicated news content highlighting moderation reduced perceived ideological polarization, which reduced affective polarization. The direct effect of the moderation condition on affective polarization in the opposite direction suggests a potential unobserved mediational pathway (Igartua & Hayes, 2021). These results also supported H2 and answered the RQ (in part), by providing support for the benefits of the polarizing content warning. When this warning was presented with the news article, it reduced perceived societal ideological polarization, which decreased affective polarization. Study 1 suggests the polarizing content warning can aid journalists in reporting on societal polarization without indirectly promoting further affective polarization. Yet, the polarizing content warning was bold and attention grabbing, is there a way to implement the warning in a more nuanced way?

## Study 2

The goal of Study 2 was to replicate Study 1 while considering other ways to implement the polarizing content warning. Given that inoculation theory suggests inoculation message effectiveness is dependent on context and messaging style (e.g., Compton & Pfau, 2005), we explored in Study 2 whether implementing warnings *within* a news article is beneficial (as compared to a warning separate from the text similar to Study 1), as a within-text warning may be even more straightforward for journalists to implement. Here, we tested a more subdued polarizing content warning within the news article itself. We tested one in-text intervention in Study 2a and two other (more obvious) iterations in Study 2b. Study 2 considers H1, H2, and the RQ.

### Study 2a sample

American participants ($N = 683$) were recruited from MTurk. Those who failed attention or manipulation checks or identified themselves as "true independents" were removed from analyses, leaving 543 participants (mean age (SD) = 42.48 (12.18), 44.20% male, 57.46% liberal).

### Study 2a procedure

Participants followed the same procedure as Study 1, though the news articles were slightly adapted. While in Study 1 we highlighted specific polarizing events (e.g., the Capitol riot), here (and in all subsequent studies), the discussion about polarization was more abstract and generalized. We did this to test whether effects would be observed with varying types of polarization news articles. Participants were randomly assigned to one of the conditions 1) polarization, 2) moderation, 3) with-text polarizing content warning (identical to Study 1), or 4) in-text polarizing content warning, where participants read the same political polarization article as those in the polarization condition; however, their article included an extra paragraph in-text discussing how findings from scientific research suggest news articles about polarization increase perceived ideological polarization and affective polarization. See page 7 of Supplementary Materials.

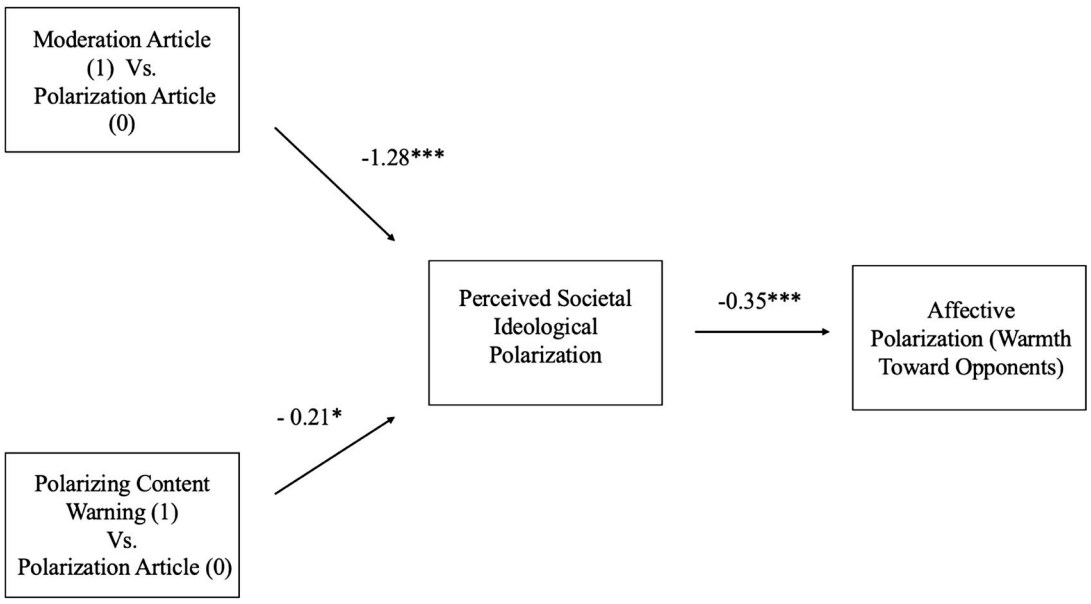


**Figure 2.** PROCESS mediation analysis (Study 1). The polarization article condition is the reference condition (entered as 0 in the model). *** indicates $p < .001$, * indicates $p < .05$.

**Table 1.** PROCESS mediation analyses for Study 1

| Outcome variable | Mediator | Effect of condition on mediator B(SE), 95% CI | Effect of mediator on DV B(SE), 95% CI | Direct effect B(SE), 95% CI | Total effect B(SE), 95% CI | Indirect effect B(SE), 95% CI |
|---|---|---|---|---|---|---|
| | | Comparing polarization article (0) to moderation article (1) | | | | |
| Affective polarization | Perceived societal ideological polarization | $B = -1.28$ (0.10) −1.49, −1.08 | $B = -0.35$ (0.06) −0.46, −0.23 | $B = -0.45$ (0.14) −0.73, −0.17 | $B = -0.008$ (0.13) −0.26, 0.24 | B = 0.44 (0.08) 0.28, 0.60 |
| | | Comparing polarization article (0) to polarizing content warning (1) | | | | |
| Affective polarization | Perceived societal ideological polarization | $B = -0.21$ (0.10) −0.42, −0.009 | $B = -0.35$ (0.06) −0.46, −0.23 | $B = -0.11$ (0.12) −0.34, 0.13 | $B = -0.03$ (0.13) −0.28, 0.21 | $B = 0.07$ (0.03) 0.02, 0.13 |

*Note*. All analyses are standardized. Affective polarization is assessed here through ratings of outgroup warmth.

### Results

Supporting H1, PROCESS mediation (Hayes & Rockwood, 2020) suggested the moderation (1) vs. polarization (0) article reduced perceptions of societal ideological polarization which decreased affective polarization, indirect effect = 0.39, SE = 0.08, 95% CI: 0.25–0.54. As in Study 1, the direct effect of the moderation condition shaped affective polarization ($B$=−0.31, SE = 0.13, 95% CI: −0.57, −0.0). Similar to Study 1, we again found that the direct effect of the moderation condition actually *increased* affective polarization in the mediation model only). Further research should examine other mediators that may explain the mechanism of how moderation news content affects affective polarization.

Supporting H2, the with-text polarizing content warning (1) vs. the polarization (0) article reduced perceived societal polarization which diminished affective polarization, indirect effect = 0.07, SE = 0.03, 95% CI: 0.02–0.13. The direct effect of the model was not significant ($B$ =−0.05, SE = 0.12, 95% CI: −0.28, 0.18). Results answered the RQ by suggesting with-text warnings are beneficial. To further answer H2, we compared the in-text polarizing content warning (1) to the polarization (0) article. The indirect effect of condition on affective polarization was not significant, indirect effect = 0.03, SE = 0.02, 95% CI: −0.02, 0.08, not supporting H2. The direct effect was also not significant ($B$=−0.14, SE = 0.12, 95% CI: −0.37, 0.10). See Figure 3 for path coefficients, Table 2 for model details, and Supplementary Table S3 for mean differences by condition.

### Study 2b

Study 2b was a follow-up to the null mediation results for the in-text warning condition in Study 2a. We recruited 505 participants on MTurk (369 of which passed attention and manipulation checks and did not identify as true independents; mean age (SD)=42.68 (12.92), 44.44% male, 57.99% liberal). We tested two new iterations of the in-text polarizing content warning condition that were potentially more obvious to readers (i.e., varying details and formatting). We hoped these changes would provide a secondary avenue for journalists to warn of the potentially ill effects of reading articles about political polarization. These new in-text iterations were also not beneficial for reducing perceived ideological polarization (and affective polarization). See page 14 of Supplementary Materials for details.

### Discussion

Study 2 provides clarity on the usability of the polarizing content warning in journalism. After testing a variety of in-text versions (which varied in terms of details and formatting), we found no effect on participants perceptions of societal polarization or their affective polarization. This may suggest that the polarizing content warning must be attention grabbing or that the warning needs to be positioned outside the text. It could be that if the warning is not obvious and attention grabbing, the threat (i.e., the first step of the inoculation process; Compton et al., 2021a), will not be recognized, making the warning ineffective. However, these are postulations and cannot be assessed by the current data. These findings support theorizing related to inoculation theory (e.g., Compton & Pfau, 2005) suggesting that inoculation message effectiveness is dependent on the style and context of the message (e.g., where it is, what it looks like, etc). Future research should disentangle when and why in-text and with-text warning labels are beneficial.

Study 2 supported H1 and H2 and Levendusky & Malhotra's (2016) results. While moderation reporting is beneficial for shifting perceptions of society's ideological polarization (which reduces affective polarization), journalists can report on societal polarization without further dividing readers by including the with-text polarizing content warning. This warning reduced participants' beliefs that society is polarized (and indirectly their affective polarization). These results answer the RQ, by suggesting the polarizing content warning is effective when presented alongside news articles but may not be effective when presented within media content.

## Study 3

While Studies 1 and 2 provided support for the benefits of the polarizing content warning in news articles on journalistic news platforms (i.e., USA Today), many access news content via social media (Shrearer, 2021). News outlets like USA Today frequently post news stories on social media platforms like Twitter to increase reach of their journalism online (e.g., Welbers & Opgenhaffen, 2018). Based on these practical realities, it seems important to consider whether the polarizing content warning is similarly effective on social media platforms. However, there are also important theoretical considerations for assessing the effectiveness of such a warning system on social media. Recent advancements of inoculation theory have pointed to how inoculation on social media (vs. traditional media sources) may in some cases lead to divergent results. For example, there are mixed results for the effectiveness of inoculation in the context of public relations on social media (e.g., Compton et al., 2021b). This suggests that

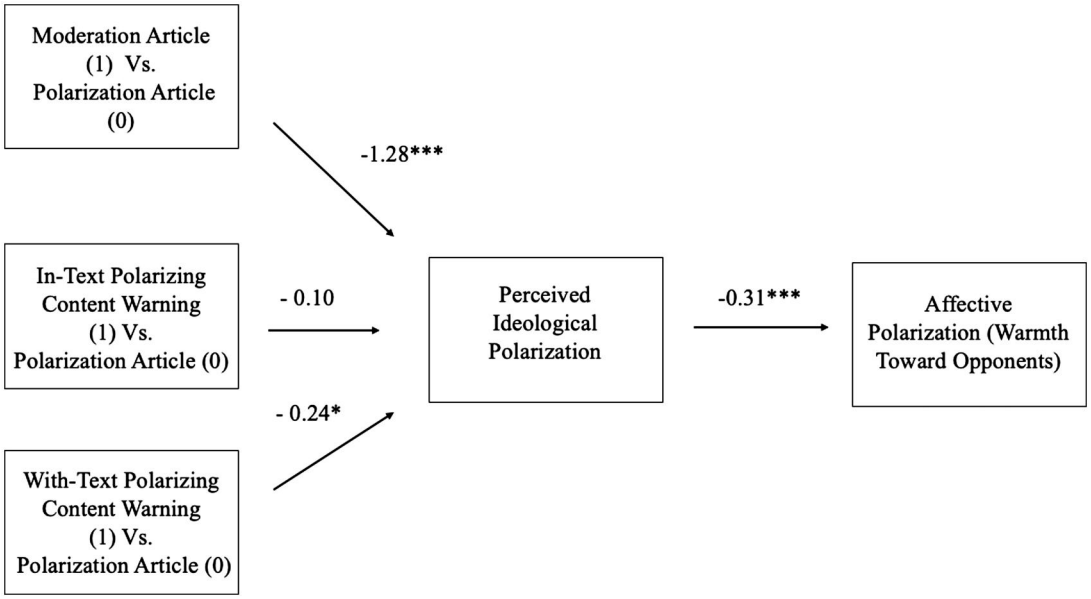


**Figure 3.** PROCESS mediation analysis (Study 2a). The polarization article condition is the reference condition (entered as 0 in the model). *** indicates $p < .001$, * indicates $p < .05$.

**Table 2.** PROCESS mediation analyses for Study 2a

| Outcome variable | Mediator | Effect of condition on mediator B(SE), 95% CI | Effect of mediator on DV B(SE), 95% CI | Direct effect B(SE), 95% CI | Total effect B(SE), 95% CI | Indirect effect B(SE), 95% CI |
|---|---|---|---|---|---|---|
| | | Comparing polarization article (0) to moderation article (1) | | | | |
| Affective Polarization | Perceived Societal Ideological Polarization | $B = -1.28$ (0.10) $-1.48, -1.08$ | $B = -0.31$ (0.05) $-0.40, -0.21$ | $B = -0.31$ (0.13) $-0.57, -0.06$ | $B = 0.08$ (0.12) $-0.15, 0.31$ | $B = 0.39$ (0.08) $0.25, 0.54$ |
| | | Comparing polarization article (0) to banner polarizing content warning (1) | | | | |
| Affective Polarization | Perceived Societal Ideological Polarization | $B = -0.24$ (0.10) $-0.44, -0.04$ | $B = -0.31$ (0.05) $-0.40, -0.21$ | $B = -0.05$ (0.12) $-0.28, 0.18$ | $B = 0.02$ (0.12) $-0.21, 0.26$ | $B = 0.07$ (0.03) $0.02, 0.13$ |
| | | Comparing polarization article (0) to in-text polarizing content warning (1) | | | | |
| Affective Polarization | Perceived Societal Ideological Polarization | $B = -0.10$ (0.11) $-0.30, 0.11$ | $B = -0.31$ (0.05) $-0.40, -0.21$ | $B = -0.14$ (0.12) $-0.37, 0.10$ | $B = -0.11$ (0.12) $-0.35, 0.14$ | $B = 0.03$ (0.02) $-0.02, 0.08$ |

*Note.* All analyses are standardized. Affective polarization is assessed here through ratings of outgroup warmth.

it is paramount to consider the effectiveness of the polarizing content warning in social media settings.

Additionally, so far, we have focused on presenting the polarizing content warning *with* news content. Inoculation theory is grounded on the idea that people are forewarned about the coming "attack" of ideas (see Lewandowsky & van der Linden, 2021). While others have also developed inoculation messages that were presented alongside content (e.g., misinformation; Brashier et al., 2020), and some scholars suggest presenting warnings alongside content is beneficial because the content is not yet fully processed (Moravec et al., 2020), it is still important to really "forewarn" people. To do this, and in line with inoculation theory, it is essential to consider the effectiveness of the polarizing content warning as a pre-bunking technique.

In Study 3, we assess these theoretical and practical considerations and test H2 and the RQ by assessing whether the political content warning can also be used in social media settings alongside news content or as a pre-bunking warning. In Study 3 and all subsequent studies, we do not include the moderation condition we are focused on assessing the robustness and effectiveness of the polarizing content warning.

### Sample

American participants ($N = 813$) were recruited from MTurk. Those who failed attention or manipulation checks or identified themselves as "true independents" were removed from analyses, ($N = 656$, mean age (SD)=40.68 (12.49), 39.33% male, 55.79% liberal).

### Procedure

Participants were randomly assigned to one of three conditions. In the polarization condition participants read a social media post from USA Today where the news site's Twitter page posted a link to a news article about polarization. This article looked similar to what actual news story links look like when posted on Twitter. See page 18 of Supplementary Materials for details. In the with-text polarizing content warning condition, participants were presented the warning alongside the news article in the polarization condition. In

the pre-bunking polarizing content warning condition, participants were first exposed to the warning and afterward saw the identical social media post to those in the polarization condition. Participants then responded to the perceived societal ideological polarization and affective polarization measures.

### Results

Supporting H2, PROCESS analyses suggested the with-text polarizing content warning (1) vs. the polarization (0) condition reduced perceptions of societal ideological polarization which reduced affective polarization, indirect effect = 0.05, SE = 0.02, 95% CI: 0.007–0.09. The direct effect was not significant ($B = 0.00$, SE = 0.09, 95% CI: −0.18, 0.18).

The pre-bunking polarizing content warning (1) vs. the polarization (0) article lessened perceived societal ideological polarization which reduced affective polarization, indirect effect = 0.09, SE=0.03, 95% CI: 0.05–0.15. The direct effect was not significant ($B = -0.03$, SE = 0.10, 95% CI: −0.22, 0.15). See Figure 4 for path coefficients, Table 3 for model details, and Supplementary Table S8.

### Discussion

The results answer the RQ by showing the polarizing content warning is effective both as a pre-bunking and with-text strategy. Additionally, Study 3 provides further support for H2 by indicating the benefits of the polarizing content warning on social media.

## Study 4

Not only have inoculation scholars argued for the importance of considering the communication context for inoculation messaging (e.g., traditional vs. social media; Compton et al., 2021b), but the *source* of these messages also may play an important role. Advancements in inoculation theory suggest that it is important to consider the source of these messages (for review see Compton & Pfau, 2005), arguably because factors like source trust and credibility are essential for whether warnings of threats and counter arguments against threats are persuasive (see Compton et al., 2021a for discussion). Importantly, though few have examined the effect of sources on inoculation effectiveness (Compton et al., 2021a), some research suggests the effectiveness of inoculation interventions is dependent on source credibility (Traberg & van der Linden, 2022) and that source credibility may be lower for news on social media as compared to traditional online news media (Besalú & Pont-Sorribes, 2021), making it imperative to examine the effectiveness of inoculation warnings from both news and social media sources.

Given calls from theoretical research to consider source effects and credibility (Compton & Pfau, 2005), and the lack of research exploring whether source effects shape the effectiveness of inoculation messaging, Study 4 explores whether the source (i.e., Twitter or USA Today) of the polarizing content warning matters. Thus, this study addresses a key theoretical gap in inoculation theory while also testing H2 and answering the RQ.

### Sample

American participants ($N = 853$) were recruited from MTurk. Individuals who failed attention and manipulation checks, or identified themselves as "true independents," were removed from analyses ($N = 621$, mean age (SD)=39.87 (12.88), 42.19% male, 57.33% liberal).

### Procedure

Like Study 3, participants read a tweet from USA Today about a news article from the media site (See Supplemental Materials page 23). Participants either saw no content warning (i.e., the polarization condition), the warning presented alongside the tweet (with-text polarizing content warning), or saw the warning before reading the article (pre-bunking polarizing content warning). In the warning conditions, participants were told the warning was from USA Today or Twitter. Participants responded to the measures used in previous studies.

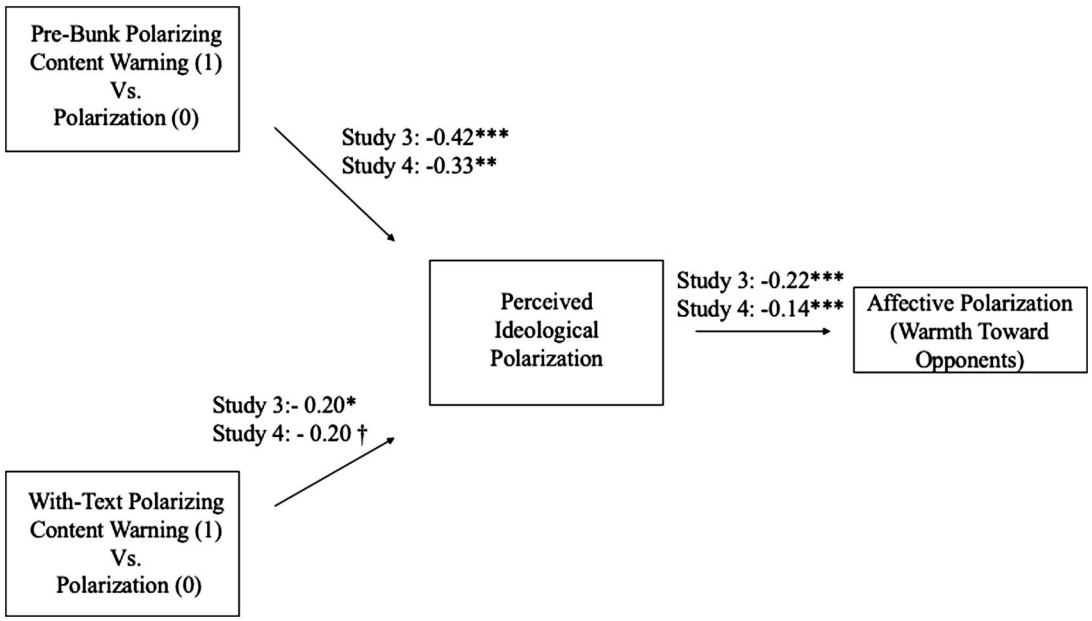


**Figure 4.** PROCESS mediation analysis (Studies 3 and 4). The polarization article condition is the reference condition (entered as 0 in the model). *** indicates $p < .001$, ** indicates $p < .01$, * indicates $p < .05$, † indicates $p < .10$.

**Table 3.** PROCESS mediation analyses for Studies 3 and 4

| Outcome variable | Mediator | Effect of condition on mediator B(SE), 95% CI | Effect of mediator on DV B(SE), 95% CI | Direct effect B(SE), 95% CI | Total effect B(SE), 95% CI | Indirect effect B(SE), 95% CI |
|---|---|---|---|---|---|---|
| **Study 3** | | | | | | |
| Comparing polarization article (0) to with-text polarizing content warning (1) | | | | | | |
| Affective polarization | Perceived societal ideological polarization | $B = -0.20$ (0.09) −0.39, −0.02 | $B = -0.22$ (0.04) −0.30, −0.14 | $B = 0.00$ (0.09) −0.18, 0.18 | $B = 0.05$ (0.10) −0.14, 0.23 | $B = 0.05$ (0.02) 0.007, 0.09 |
| Comparing polarization article (0) to pre-bunking polarizing content warning (1) | | | | | | |
| Affective polarization | Perceived societal ideological polarization | $B = -0.42$ (0.09) −0.60, −0.23 | $B = -0.22$ (0.04) −0.30, −0.14 | $B = -0.03$ (0.10) −0.22, 0.15 | $B = 0.06$ (0.10) −0.13, 0.25 | $B = 0.09$ (0.03) 0.05, 0.15 |
| **Study 4** | | | | | | |
| Comparing polarization article (0) to with-text polarizing content warning (1) | | | | | | |
| Affective polarization | Perceived societal ideological polarization | $B = -0.20$ (0.11) −0.41, 0.01 | $B = -0.14$ (0.04) −0.22, −0.06 | $B = 0.05$ (0.11) −0.15, 0.26 | $B = 0.08$ (0.11) −0.13, 0.29 | $B = 0.03$ (0.02) 0.0004, 0.07 |
| Comparing polarization article (0) to pre-bunking polarizing content warning (1) | | | | | | |
| Affective polarization | Perceived societal ideological polarization | $B = -0.33$ (0.10) −0.53, −0.13 | $B = -0.14$ (0.04) −0.22, −0.06 | $B = 0.19$ (0.10) −0.01, 0.39 | $B = 0.24$ (0.10) 0.03, 0.44 | $B = 0.05$ (0.02) 0.01, 0.10 |

*Note.* All analyses are standardized. Affective polarization is assessed here through ratings of outgroup warmth.

### Results

Results indicated that it did not matter whether a warning was from Twitter or USA Today. There were no significant differences between sources on both perceived ideological polarization and affective polarization. See page 26 of Supplementary Materials. Given the lack of source effects, we collapsed conditions (a pre-registered decision). We collapsed both with-text warning conditions and both pre-bunking warning conditions. See Supplementary Table S11 for mean differences by condition.

We assessed H2 and RQ with PROCESS mediational analysis which suggested the pre-bunking warning (1) vs. the polarization (0) article reduced perceptions of societal ideological polarization which subsequently reduced affective polarization, indirect effect = 0.05, SE = 0.02, 95% CI: 0.01–0.10. The direct effect was not significant ($B = 0.19$, SE = 0.10, 95% CI: −0.01, 0.39).

Analyses revealed the with-text polarizing content warning (1) vs. the polarization (0) condition, reduced perceptions of societal ideological polarization (trending, $p$=.062), which diminished affective polarization. Importantly, in line with the other results, the indirect effect was significant ($B = 0.03$, SE = 0.02, 95% CI: 0.0004–0.07). The direct effect was not significant ($B = 0.05$, SE = 0.11, 95% CI: −0.15, 0.26). See Figure 4 for path coefficients and Table 3 for model details.

### Discussion

Study 4 extended the findings of Study 3 and provided insights for inoculation theory by considering the source of warnings. Results suggested the source of the warning did not matter. Study 4 supported H2 by suggesting the polarizing content warning reduces perceived societal polarization and indirectly reduces affective polarization. This study reinforced answers to the RQ by reiterating the benefits of the pre-bunking and with-text warning on social media.

## Study 5

The previous studies point to the benefits of the polarizing content warning for both online news articles and social media. However, are there ways to implement this warning into real platforms? Research related to inoculation theory emphasizes the importance of testing inoculation messages in externally valid settings (e.g., Roozenbeek et al., 2022). This suggests it is essential to test whether such warnings are effective within the constraints of warning systems currently used by social media platforms (e.g., to combat misinformation). Further, many journalists and news outlets may be unaware of the polarizing consequences of their news reports regarding societal polarization—meaning they may not think to use the polarizing content warning when it could be helpful. Additionally, some news outlets (e.g., partisan news outlets) may be unwilling to include warnings as it is part of their business model to provide polarizing and attention-grabbing news to their audiences (e.g., Levendusky, 2013; Munger, 2020).

Based on these theoretical and practical considerations, in Study 5 we explored whether the polarizing content warning could be adapted to Twitter and Facebook's current warning systems (for combating misinformation) to test whether platforms can warn users when journalists are unable or unwilling to do so. Study 5 tests H2 and answered the RQ.

### Sample

American participants ($N = 1{,}026$) were recruited from MTurk. Individuals who failed attention and manipulation checks or identified themselves as "true independents" were removed from analyses ($N = 843$, mean age (SD) = 39.85 (12.82), 38.79% male, 55.04% liberal)[4].

### Procedure

Participants were randomly assigned to one of the following conditions: 1) Facebook polarization, 2) Facebook pre-bunking polarizing content warning, 3) Twitter polarization, or 4) Twitter pre-bunking polarizing content warning. In the polarization conditions, participants read a post from USA Today about a news article they published related to polarization on the platform. In the pre-bunking conditions, participants saw the polarized content warning which was adapted

to be comparable to Facebook and Twitters' real pop-up misinformation warnings, and then saw the same post from USA Today as those in the polarization condition. See page 30 of Supplementary Materials. Participants responded to the same measures from previous studies.

### Results

Analyses revealed no significant differences by platform. Based on these results, we collapsed across social media platforms and compared all polarization conditions to all polarizing content warning conditions (i.e., two conditions). PROCESS mediational analyses revealed the pre-bunking polarizing content warning (1) vs. the polarization condition (0), reduced perceptions of societal ideological polarization, which reduced affective polarization, indirect effect = 0.04, SE = 0.01, 95% CI: 0.02–0.07. The direct effect of the model was not significant ($B = -0.01$, SE = 0.07, 95% CI: −0.15, 0.12). See Figure 5 for path coefficients, Table 4 for model details, and Supplementary Tables S13 and S14 for mean differences[5].

### Discussion

Study 5 provides insight into how to implement the polarizing content warning within existing social media warning systems. The warning reduced how polarized people think society is, which reduced affective polarization (supporting H2). Study 5 also answers the RQ by highlighting the benefits of pre-bunking in social media settings. These results indicate polarizing content warnings on social media do not necessarily have to be used by journalists or news outlets but that platforms can also warn social media users about online journalism that may be polarizing. We find such warnings indirectly prevent audiences from becoming more affectively polarized.

## General discussion

The polarizing content warning provides a strategy for journalists (and social media sites) to report and share news about growing levels of political polarization while indirectly reducing affective polarization. This research provides important insights into how the media can *reduce* affective polarization, and corroborates the well-supported connection between *perceptions* of societal polarization and actual affective polarization (e.g., Druckman et al., 2022; Enders & Armaly, 2019; Moore-Berg et al., 2020).

This warning is effective as both a pre-bunking and with-text warning system on both online news and social media sites. While this warning is not effective with subtle in-text warning cues (Study 2), it does not matter who the source of the warning is (i.e., when comparing USA Today vs. Twitter), and it can be applied to current warning systems already used to combat misinformation by leading social media platforms (i.e., Twitter and Facebook). Importantly, presenting news content about societal polarization alongside the polarizing content warning, indirectly reduces affective polarization as compared to presenting the news content about societal polarization alone. This means journalists can still share the same content but reduce partisan animus among readers.

### The polarizing content warning and trust in media

While it is fruitful to examine the potential benefits of the polarizing content warning, it is also important to examine whether this warning negatively affects peoples' attitudes toward the news media platform itself. If it does, it would likely mean journalists and media companies would be de-incentivized to use such a warning system (in order to maintain trust with their readers). Exploratory analyses revealed that in Studies 2–5 the polarizing content warning did not

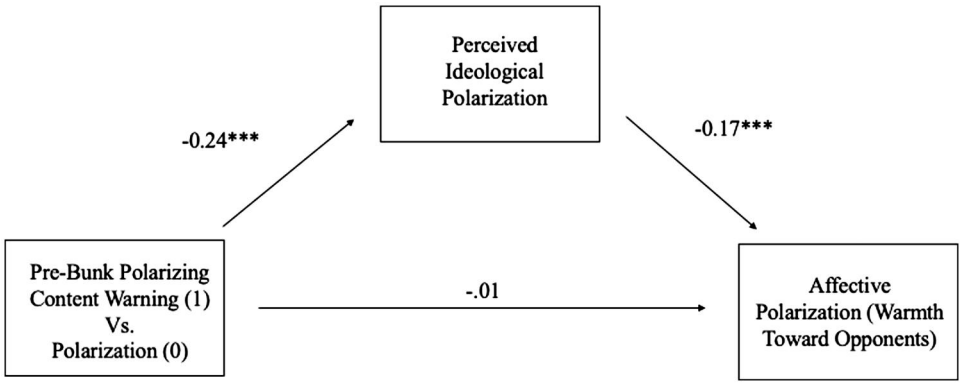


**Figure 5.** PROCESS mediation analysis (Study 5). The polarization article condition is the reference condition (entered as 0 in the model). *** indicates $p < .001$. The estimate on the c pathway between condition and affective polarization represents the direct effect.

**Table 4.** PROCESS mediation analyses for Study 5

| Outcome variable | Mediator | Effect of condition on mediator B(SE), 95% CI | Effect of mediator on DV B(SE), 95% CI | Direct effect B(SE), 95% CI | Total effect B(SE), 95% CI | Indirect effect B(SE), 95% CI |
|---|---|---|---|---|---|---|
| Comparing polarization (0) to pre-bunking polarizing content warning (1) | | | | | | |
| Affective polarization | Perceived societal ideological polarization | $B = -0.24$ (0.07)<br>−0.38, −0.11 | $B = -0.17$ (0.03)<br>−0.24, −0.11 | $B = -0.01$ (0.07)<br>−0.15, 0.12 | $B = 0.03$ (0.07)<br>−0.11, 0.16 | $B = 0.04$ (0.01)<br>0.02, 0.07 |

*Note*. All analyses are standardized. Affective polarization is assessed here through ratings of outgroup warmth.

increase distrust for the news source as compared to conditions without the warning. This suggests the warning may be a useful tool for journalists that will also not affect trust toward media sources.

However, in Study 1, the polarizing content warning led to significantly less trust in *USA Today* as compared to the polarization article ($p = .01$), see page 40 of Supplementary Materials. While it is unclear why this inconsistent finding occurred, we posit it may have been related to the more specific discussions of polarizing events in the United States (e.g., the Capitol riot) in the news articles in Study 1. In all other studies, the news articles had more abstract discussions about polarization trends. It could be that discussions about specific extreme and polarizing events in combination with the polarizing content warning reduced trust in the media source in Study 1—an assumption supported by previous research on fact checking in news media (Li et al., 2022). Unfortunately, this possibility cannot be examined with the current data.

Future research should assess under which circumstances the polarizing content warning affects trust in the news media source. However, we do want to point out that finding no effect in 5 of the 6 samples does provide some promising evidence of the warning being a powerful tool for journalistic communication and human communication research more broadly.

## Implications for inoculation theory and research

This research advances inoculation theory (McGuire, 1964) in a variety of ways. First, we apply inoculation theory into a new realm of research by focusing on finding ways to reduce polarization. Here, we examine how inoculation messaging can be used in media settings to reduce perceived societal ideological polarization and affective polarization, through both highlighting the threat of polarization and counterarguing against it (in line with previous inoculation research; Compton et al., 2021a). Further, we suggest the knowledge gained from this research can theoretically extend inoculation theory as we have identified a potentially relevant mechanism for understanding the effectiveness of inoculation—reductions in perceptions of society as polarized. We hope this work theoretically informs further research on inoculation more generally and provides a first step for scholars to understand how we can inoculate in contexts of polarization.

This research emphasizes how we can inoculate people against content known to make people see society as more ideologically divided (e.g., news articles about polarization; Levendusky & Malhotra, 2016). However, we note that while we presume that the polarized content warning reduces perceived ideological polarization through some form of counterarguing, we do not explicitly measure this process in our studies—an important direction for future research. Prior inoculation research suggests counterarguing can be assessed through thought-listing techniques (Petty et al., 1976), where people list counter arguments (Parker et al., 2012). In our case, this could be listing arguments for why societal ideological polarization may actually be less than they perceive it to be. We encourage future research explicitly examine the elements of inoculation (i.e., warning of threats and counterarguing against them; Compton et al., 2021a) as underlying mechanisms for explaining the effectiveness of warnings like the polarizing content warning.

Additionally, through testing the polarizing content warning in a variety of contexts (news [Studies 1 & 2] and social media [Studies 3–5]), with a variety of messaging styles (Study 2a and 2b), with different message sources (Studies 4 and 5), and with warnings presented beforehand and with news content (Studies 3–5), we examine many of the key factors inoculation theory argues can influence the effectiveness of inoculation messages (see Compton & Pfau, 2005). Overall, we find that the polarizing content warning is effective in news and social media, can come from news or social media platforms, and can be presented beforehand or with news content, but it must be attention grabbing and alongside new content rather than within the text.

## Implications for journalism and news on social media

While the current research is narrow in its focus (i.e., focusing on whether a warning system is helpful for reducing the polarizing effects of media content discussing polarization), it opens many avenues for future research and theoretical and ethical discussions on how these kinds of warning systems can be applied in journalism and on social media sites. For example, it may be possible to include such a warning label with Tweets that include polarizing messaging or with news stories about political opponents, both of which have been connected to increasing partisan animosity among media users (Kubin & von Sikorski, 2021).

Another important consideration for this work is that individual journalists cannot always be made responsible to deal with the consequences of their fact-based reporting. However, they should be aware of the unintended effects certain types of reporting can have. In addition, journalistic media organizations and other institutions should further grapple with these issues and develop guidelines and procedures for media providers to follow. In the current research, we demonstrate that warnings work effectively. However, when or if these warnings are used should be thoroughly considered and be accompanied by future research.

We also want to make note of the potential differences in application of such a warning system in news and social media contexts. Arguably, the polarizing content warning could be implemented in social media platforms in similar ways to established misinformation warnings (e.g., flagging techniques; Kim et al., 2018). Applying such a warning to news media may be a bit more challenging. Questions remain regarding whether journalists would be willing to include such warning systems and how to assess these warning systems in real news media coverage. However, news organizations have worked alongside researchers to find ways to reduce political polarization previously (see Darr et al., 2021), suggesting some news organizations are motivated to find ways to heal divisions. We encourage scholars to work alongside journalists to further explore the applicability of this intervention in real media content.

Additionally, questions remain regarding whether mass implementation of these warning systems could lead users of media content to habituate to such warning systems, making the warning less effective—a common concern related to misinformation flags (e.g., Vance et al., 2019). Furthermore, we only explore whether the warning was beneficial at one time point and did not assess whether effects remain stable over time. We encourage future research to assess such possibilities of habituation and explore whether there are long-term benefits for using the polarizing content warning.

This research also addresses a key limitation of Levendusky and Malhotra's (2016) moderation intervention. While we were able to conceptually replicate their original findings (i.e., the moderation approach being very effective), journalists still need to report on the realities of polarization in polarized democracies like the United States. The polarizing content warning addresses this limitation and is a first step in solving the dilemma (i.e., not reporting vs. further increasing polarization) by providing one strategy that can allow journalists to report on polarization without indirectly affectively polarizing readers further.

## Limitations and future directions

While there are many practical and theoretical contributions of this work, there are also limitations. All studies relied on convenience samples from MTurk. While MTurk has been cited as having comparable data quality to other participant pools (Kees et al., 2017; McEwan, 2020), future research should test the benefits of the polarizing content warning with other samples (e.g., quota-based samples). Relatedly, we did not consider whether potentially relevant demographics (e.g., education level) could influence the effectiveness of the polarizing content warning—a consideration that should be of focus in future research. Additionally, the news source in all studies was *USA Today*, a relatively mainstream and moderate news source (Pew Research Center, 2014). Therefore, it is unclear whether effects would replicate with more partisan news sources (e.g., Fox News). Further, like Levendusky and Malhotra (2016), we focused solely on the American context, and do not make postulations that effects necessarily generalize to other societies or situations.

The goal of the current research was to explore, as a first step, whether such a warning system could be effective in indirectly reducing affective polarization. We believe this work is a promising first step in understanding when and where we can implement such warning systems, but many questions remain. For example, (when) are journalists willing to use such a warning system? And who should make decisions on where and when it is appropriate to use this warning system in media? We also recommend future research consider adaptations to the warning itself (e.g., adaptations to the language used in the warning), to explore whether these adaptations are more or less effective than the one proposed in this initial research.

Additionally, all experiments relied on self-report measures. Future research should consider other ways to assess affective polarization (e.g., behavioral approaches). These experiments also did not consider participants' trust (or belief) in science, which are factors that could be potentially relevant given the warning mentions scientific research. While we believe finding an effect of the polarizing content warning across participants highlights its efficacy, it may be beneficial to consider how individual differences shape the effectiveness of such warning systems.

We also want to note caution when interpreting the indirect effects on affective polarization. While previous research suggests such indirect (mediation) effects are meaningful (Igartua & Hayes, 2021), the effect of perceived societal ideological polarization on affective polarization is correlational in nature. This should be taken into account when interpreting the results. Although the indirect effects on (reduced) affective polarization are in line with our theorizing and previous empirical findings, future research should test whether manipulating levels of perceived societal ideological polarization reduces affective polarization and if warning systems in other contexts can (indirectly) reduce affective polarization.

Some readers may wonder the value of warning about the potentially unintended polarizing consequences of news content if the total effect of the polarizing content warning on affective polarization is not significant (even though the indirect effect is significant). Zhao et al. (2010) have pointed to how these patterns can occur for a variety of reasons, such as when there are multiple mediators working in opposite directions but argue that such indirect effects are still statistically meaningful. These findings may suggest that other underlying mechanisms, not tested here, may also mediate (perhaps in opposite directions) effects—dampening the total effect. To further understand the normative implications and effectiveness of this intervention, scholars should explore mediators beyond perceived ideological polarization.

Further, while the polarizing content warning did not directly reduce affective polarization, it did directly reduce perceived ideological polarization. Intervention research is frequently focused on helping people recognize ideological differences are not as extreme as they believe (e.g., Lees & Cikara, 2020; Voelkel et al., 2023) and reducing perceived ideological polarization to combat negative affective evaluations of opponents (Enders & Armaly, 2019)—all processes that can contribute to a healthier democracy (Overgaard et al., 2022). This suggests reducing perceptions of society being polarized may have practical value in itself.

Relatedly, although we believe that our stimuli were generally perceived as authentic, future research should explicitly test this. Specifically, it should be analyzed if warnings affect credibility and readability of the article. Our data show no general brand-damaging effects (i.e., assessment of the news source USA Today) triggered by the warning. However, future research should test this more specifically, also considering that warnings may improve perceived article quality, as warnings may also be interpreted as journalistic quality indicators.

Additionally, effects were robust, but small, which is not unusual for media effects research and findings in related fields (see Valkenburg & Peter, 2013). Although the warning system is indeed effective, readers should not overinterpret the results of this intervention. Further, based on our study design (where we asked participants to rate warmth toward Republicans and Democrats—a standard practice in past research; Druckman et al., 2022; Levendusky & Malhotra, 2016), we excluded people who were political independents from our study design. Future research should assess how these warnings affect non-partisans.

This research tested the effectiveness of the polarizing content warning in a variety of contexts (e.g., pre-bunking vs. with-text warnings, source effects, news vs. social media) and thus did not focus on examining other underlying mechanisms at play. It could be that our warning system may actually bridge inoculation theory and other theoretical research in communication. For example, warnings about perceived societal polarization may also be relevant for future research informed by the "third person effect" (Perloff, 1993) and the Influence of Presumed Media Influence (IPMI) theory (Gunther & Storey, 2003; see also Tal-Or et al., 2010). Warnings implemented in media may be a relevant cue for people that other individuals exposed to the same news

article are likely to be affected by the message (article about polarization in the United States). The warning may then trigger the third person effect (i.e., beliefs that other people [but not yourself] are affected by the article [and that is why the media company uses the warning]), which in turn shapes perceptions of societal polarization (or beliefs about the extent to which others perceive societal polarization). Future research might test this (and other) additional theoretical paths and underlying mechanisms.

Furthermore, given that the current research explores perceptions of societal ideological polarization (i.e., how polarized I think society is), misperceived polarization (e.g., Lees & Cikara, 2021) is a highly relevant construct for understanding whether media content about polarization (and the polarizing content warning) drive misperceptions of polarization. It could be that media content about polarization leads people to have overexaggerated perceptions of how polarized society is or that the polarizing content warning leads people to have underexaggerated perceptions of societal polarization. In the current research, we do not consider whether peoples' perceptions of societal polarization are accurate. Finally, we would like to emphasize that this comprehensive presentation of the limitations of the present studies should also be understood as a suggestion for future research. We hope that others will conduct new studies based on these initial results.

## Conclusion

Reporting on polarization in society is paramount for journalists but may contribute to further polarization (Levendusky & Malhotra, 2016). We find that this is because such reporting shifts how polarized people believe society is, which further enflames affective polarization. The polarizing content warning provides a new strategy for journalists to report on political polarization while still healing political divisions.

## Supplementary material

Supplementary material is available at *Human Communication Research* online.

## Data availability

Data, materials, and pre-registrations for all studies are available at: https://osf.io/na3dy/?view_only=c970731583d6439baf4ade59bd033dda.

## Funding and Disclosure statements

We are grateful for support from the Schlieper Foundation to Christian von Sikorski (No. 4507-LD-28202).

*Conflicts of interest*: None declared.

## Open science framework badges

**Open Materials**
The components of the research methodology needed to reproduce the reported procedure and analysis are publicly available for this article.

**Open Data**
Digitally shareable data necessary to reproduce the reported results are publicly available for this article.

**Preregistered**
Research design was preregistered.

## Notes

1. Participant party affiliation was determined using a two-step process. First, participants were asked if they identify with "the Republican Party," "the Democratic Party," or "Neither." Participants who responded with "Neither" were then asked whether they leaned more towards the Republican or Democratic party or if they "do not lean towards either party". Participants who reported not leaning toward either party were categorized as "true independents" and removed from analyses.
2. Participants were removed from analyses for failing one of our attention checks *or* for reporting being independent. Across samples, approximately 10–12% reported being independent (except in Study 2b where it was approximately 14%). These figures are in line with current estimates of independents in the American electorate (e.g., Theiss-Moore & Wagner, 2022).
3. All PROCESS mediation analyses in all studies used 5,000 iterations of bootstrapping and were conducted by using Model 4 from PROCESS macro.
4. We originally pre-registered the Facebook and Twitter conditions as separate studies (all run in the same batch). However, after data collection we recognized effects were smaller than predicted. Given that there were no significant differences between Facebook and Twitter conditions, we collapsed across conditions to gain greater power.
5. In Study 5, we also conducted analyses in SEM to assess the fit of the model. Results indicated the model was fully saturated. Further, a likelihood ratio test indicated this model had better fit than the null model $\chi^2 = 12.51$, df = 1, $p < .001$. We additionally explored whether similar effects emerged when the mediator (perceived polarization) and dependent variable (affective polarization) were switched. Results were non-significant (indirect effect = –.005, $SE = .01$, 95% CI: −0.41–0.68). These analyses were not pre-registered. See page 38 in Supplemental Materials.

## References

Ahler, D. J., & Sood, G. (2018). The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics*, *80*(3), 964–981. https://doi.org/10.1086/697253

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs*, *77*(3), 281–311. https://doi.org/10.1080/03637751003758193

Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, *3*(1), 2, 1–9. https://doi.org/10.5334/joc.91

Besalú, R., & Pont-Sorribes, C. (2021). Credibility of digital political news in Spain: comparison between traditional media and social media. *Social Sciences*, *10*(5), 170. https://doi.org/10.3390/socsci10050170

Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2020). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, *118*(5), 1–3. https://doi.org/10.1073/pnas.202004311

Clark, J. K., & Evans, A. T. (2014). Source credibility and persuasion: The role of message position in self-validation. *Personality and Social Psychology Bulleting*, *40*(8), 1024–1036. https://doi.org/10.1177/0146167214534733

Compton, J. (2013). Inoculation theory. In J. P. Dillard & L. Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice* (2nd ed., pp. 220–236). SAGE Publications. https://doi.org/10.4135/9781452218410

Compton, J., Jackson, B., & Dimmock, J. A. (2016). Persuading others to avoid persuasion: Inoculation theory and resistant health

attitudes. *Frontiers in Psychology*, 7. https://doi.org/10.3389/fpsyg.2016.00122

Compton, J. A., & Pfau, M. (2005). Inoculation theory of resistance to influence at maturity: Recent progress in theory development and application and suggestions for future research. *Annals of the International Communication Association*, *29*(1), 97–146. https://doi.org/10.1080/23808985.2005.11679045

Compton, J., van der Linden, S., Cook, J., & Basol, M. (2021a). Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories. *Social and Personality Psychology Compass*, *15*(6), e12602. https://doi.org/10.1111/spc3.12602

Compton, J., Wigley, S., & Samoilenko, S. A. (2021b). Inoculation theory and public relations. *Public Relations Review*, *47*(5), 102116. https://doi.org/10.1016/j.pubrev.2021.102116

Darr, J. P., Hitt, M. P., & Dunaway, J. L. (2021). Home style opinion: How local newspapers can slow polarization. *Public Opinion Quarterly*, *85*(4), 1155–1157. https://doi.org/10.1017/9781108950930

DeAndrea, D. C., & Holbert, R. L. (2017). Increasing clarity where it is needed most: Articulating and evaluating theoretical contributions. *Annals of the International Communication Association*, *41*(2), 168–180. https://doi.org/10.1080/23808985.2017.1304163

Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2022). (Mis)estimating affective polarization. *The Journal of Politics*, *84*(2), 1106–1117. https://doi.org/10.1086/715603

Enders, A. M., & Armaly, M. T. (2019). The differential effects of actual and perceived polarization. *Political Behavior*, *41*(3), 815–839. https://doi.org/10.1007/s11109-018-9476-2

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., … Druckman, J. N. (2020). Political sectarianism in America: A poisonous cocktail of othering, aversion, and moralization poses a threat to democracy. *Science*, *370*(6516), 553–536. doi: 10.1126/science.abe1715

Fiorina, M. (2016). Has the American public polarized? Hoover Institution. Available at: https://www.hoover.org/sites/default/files/research/docs/fiorina_finalfile_0.pdf

Fiorina, M., & Abrams, S. (2008). Political polarization in the American public. *Annual Review of Political Science*, *11*, 563–588. https://doi.org/10.1146/annurev.polisci.11.053106.153836

Fiorina, M., Abrams, S., & Pope, J. (2005). *Culture war? The myth of a polarized America*. Pearson-Longman.

Garrett, R. K., Gvirsman, S. D., Johnson, B. K., Tsfati, Y., Neo, R., & Dal, A. (2014). Implications of pro- and counterattitudinal information exposure for affective polarization. *Human Communication Research*, *40*(3), 309–332. https://doi.org/10.1111/hcre.12028

Gunther, A. C., & Storey, J. D. (2003). The influence of presumed influence. *Journal of Communication*, *53*(2), 199–215. https://doi.org/10.1111/j.1460-2466.2003.tb02586.x

Han, J., & Federico, C. M. (2018). The polarizing effect of news framing: Comparing the mediating roles of motivated reasoning, self-stereotyping, and intergroup animus. *Journal of Communication*, *68*(4), 685–711. https://doi.org/10.1093/joc/jqy025

Hayes, A. F., & Rockwood, N. J. (2020). Conditional process analysis: Concepts, computation, and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist*, *64*(1), 19–54. https://doi.org/10.1177/0002764219859633

Heesacker, M., Petty, R. E., & Cacioppo, J. T. (1983). Field dependence and attitude change: Source credibility can alter persuasion by altering message-relevant thinking. *Journal of Personality*, *51*(4), 653–666. https://doi.org/10.1111/j.1467-6494.1983.tb00872.x

Hosking, G. (2019). The decline of trust in government. In M. Sasaki (Ed.), *Trust in contemporary society* (pp. 77–103). Brill. https://doi.org/10.1163/9789004390430_007

Igartua, J., & Hayes, A. F. (2021). Mediation, moderation, and conditional process analysis: Concepts, computations, and some common confusions. *The Spanish Journal of Psychology*, *24*:e5, 1–23. https://doi.org/10.1017/SJP.2021.46

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, *76*(3), 405–431. https://doi.org/10.1093/poq/nfs038

Kalmoe, N. P., & Mason, L. (2022). *Radical American partisanship: Mapping violent hostility, its causes, & what it means for democracy*. University of Chicago Press. https://doi.org/10.7208/Chicago/9780226820279.001.0001

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An analysis of data quality: Professional panels, student subject pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*(1), 141–155. https://doi.org/10.1080/00913367.2016.1269304

Kim, J., Tabibian, B., Oh, A., Schölkopf, B., & Gomez-Rodriguez, M. (2018, February). Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In *Proceedings of the eleventh ACM international conference on web search and data mining* (pp. 324–332). https://doi.org/10.1145/3159652.3159734

Kim, Y., & Zhou, S. (2020). The effects of political conflict news frame on political polarization: A social identity approach. *International Journal of Communication*, *14*, 937–958.

Kreiss, D., & McGregor, S. C. (2023). A review and provocation: On polarization and platforms. *New Media & Society*, *26*(1), 1–24. https://doi.org/10.1177/14614448231161880

Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association*, *45*(3), 188–206. https://doi.org/10.1080/23808985.2021.1976070

Kubin, E., & von Sikorski, C. (2023). The complex relationship between media and political polarization: Understanding how the media can affectively (de)polarize citizens. *International Journal of Communication*, *17*, 5207–5222.

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behavior*, *4*, 279–286. https://doi.org/10.1038/s41562-019-0766-4

Lees, J., & Cikara, M. (2021). Understanding and combating misperceived polarization. *Philosophical Transactions of the Royal Society, B*, *376*(1822), 20200143. https://doi.org/10.1098/rstb.2020.0143

Levendusky, M. (2013). Why do partisan media polarize viewers? *American Journal of Political Science*, *57*(3), 611–623. https://doi.org/10.1111/ajps.12008

Levendusky, M., & Malhotra, N. (2016). Does media coverage of partisan polarization affect political attitudes? *Political Communication*, *33*(2), 283–301. https://doi.org/10.1080/10584609.2015.1038455

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. https://doi.org/10.1080/10463283.2021.1876983

Li, J., Foley, J. M., Dumdum, O., & Wagner, M. W. (2022) The power of a genre: Political news presented as fact-checking increases accurate belief updating and hostile media perceptions. *Mass Communication and Society*, *25*(2), 282–307, https://doi.org/10.1080/15205436.2021.1924382

Matthes, J., Knoll, J., & von Sikorski, C. (2018). "Spiral of Silence" revisited: A meta-analysis on the relationship between perceptions of opinion support and political opinion expression. *Communication Research*, *45*(1), 3–33. https://doi.org/10.1177/0093650217745429

McEwan, B. (2020). Sampling and validity. *Annals of the International Communication Association*, *44*(3), 235–247. https://doi.org/10.1080/23808985.2020.1792793

McGuire, W. J. (1964). Inducing resistance to persuasion: Some contemporary approaches. *Advances in Experimental Social Psychology*, 1, 191–229. (Reprinted in *Self and society: An anthology of readings*, pp. 192–230, by C. C. Haaland and W. A. Kaebler, Eds., 1981, Ginn Custom Publishing).

Moore-Berg, S. L., Hameiri, B., & Bruneau, E. G. (2022). Empathy, dehumanization, and misperceptions: A media intervention humanizes migrants and increases empathy for their plight but only if

misinformation about migrants is also corrected. *Social Psychological and Personality Science*, *13*(2), 645–655. https://doi.org/10.1177/19485506211012793

Moore-Berg, S. L., Hameiri, B., & Bruneau, E. (2020). The prime psychological suspects of toxic political polarization. *Current Opinion in Behavioral Sciences*, *34*, 1–6. https://doi.org/10.1016/j.cobeha.2020.05.001

Moravec, P. L., Kim, A., & Dennis, A. R. (2020). Appealing to sense and sensibility: System 1 and System 2 interventions for fake news on social media. *Information Systems Research*, *31*(3), 987–1006. doi:10.1287/isre.2020.0927

Munger, K. (2020). All the news that's fit to click: The economics of clickbait media. *Political Communication*, *37*(3), 376–397. https://doi.org/10.1080/10584609.2019.1687626

Noelle-Neumann, E. (1974). The spiral of silence a theory of public opinion. *Journal of Communication*, *24*(2), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

O'Rourke, H. P., & MacKinnon, D. P. (2018). Reasons for testing mediation in the absence of an intervention effect: A research imperative in prevention and intervention research. *Journal of Studies on Alcohol and Drugs*, *79*(2), 171–181. https://doi.org/10.15288/jsad.2018.79.171

Overgaard, C. S. B., Masullo, G. M., Duchovnay, M., & Moore, C. (2022). Theorizing connective democracy: A new way to bridge political divides. *Mass Communication and Society*, *25*(6), 861–885. https://doi.org/10.1080/15205436.2022.2119870

Perloff, R. M. (1993). Third-person effect research 1983-1992: A review and synthesis. *International Journal of Public Opinion Research*, *5*(2), 167–184. https://doi.org/10.1093/ijpor/5.2.167

Pew Research Center (2017). The partisan divide on political values grows even wider. Retrieved May 28, 2021 from: https://www.pewresearch.org/politics/2017/10/05/the-partisan divide-on-political-values-grows-even-wider/

Parker, K. A., Ivanov, B., & Compton, J. (2012). Inoculation's efficacy with young adults' risky behaviors: can inoculation confer cross-protection over related but untreated issues? *Health Communication*, *27*(3), 223–233. https://doi.org/10.1080/10410236.2011.575541

Petty, R. E, Wells, G. L., & Brock, T. C. (1976). Distraction can enhance or reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology*, *34*(5), 874–884. https://doi.org/10.1037/0022-3514.34.5.874

Pew Research Center (2014). Political Polarization and Media Habits. Retrieved July 6, 2022 from https://www.pewresearch.org/journalism/2014/10/21/political-polarization-media-habits/

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Robison, J., & Mullinix, K. J. (2016). Elite polarization and public opinion: How polarization is communicated and its effects. *Political Communication*, *33*(2), 261–282. https://doi.org/10.1080/10584609.2015.1055526

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Commun*, *5*(65), 1–10. https://doi.org/10.1057/s41599-019-0279-9

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.

Ruggeri, K., Većkalov, B., Bojanić L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., … Folke, T. (2021). The general fault in our fault lines. *Nature Human Behavior*, *5*, 1369–1380. https://doi.org/10.1038/s41562-021-01092-x

Saleem, M., Prot, S., Anderson, C. A., & Lemieux, A. F. (2015). Exposure to Muslims in media and support for public policies. *Communication Research*, *44*(6), 841–869. https://doi.org/10.1177/0093650215619214

Shrearer, E. (2021). More than eight-in-ten Americans get news from digital devices. *Pew Research Center*. Retrieved May 19 2022. https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/

Stavrakakis, Y. (2018). Paradoxes of polarization: Democracy's inherent division and the (anti-) populist challenge. *American Behavioral Scientist*, *62*(1), 43–58. https://doi.org/10.1177/0002764218756924

Tal-Or, N., Cohen, J., Tsfati, Y., & Gunther, A. C. (2010). Testing Causal Direction in the Influence of Presumed Media Influence. *Communication Research*, *37*(6), 801–824. https://doi.org/10.1177/0093650210362684

Theiss-Moore, E. A., & Wagner, M. W. (2022). *Political behavior of the American electorate* (15th edn). CQ Press.

Traberg, C. S., & van der Linden, S. (2022). Birds of a feather are persuaded together: Perceived source credibility mediates the effect of political bias on misinformation susceptibility. *Personality and Individual Differences*, *185*, 111269. https://doi.org/10.1016/j.paid.2021.111269

Valkenburg, P. M., & Peter, J. (2013). Five challenges for the future of media-effects research. *International Journal of Communication*, *7*, 197–215.

van der Linden, S., Lesierowitz, A. A., Feinberg, G. D., & Maibach, E. W. (2015). The scientific consensus on climate change as a gateway belief: Experimental evidence. *Plos One*, *10*(2), e0118489. https://doi.org/10.1371/journal.pone.0118489

van der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating against fake news about COVID-19. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.566790

Vance, A., Eargle, D., Jenkins, J. L., Kirwan, C. B., & Anderson, B. B. (2019). The fog of warnings: how non-essential notifications blur with security warnings. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)* (pp. 407–420). https://www.usenix.org/system/files/soups2019-vance.pdf

Voelkel, J. G., Stagnaro, M., Chu, J., Pink, S., Mernyk, J., Redekopp, C., Ghezae, I., Cashman, M., Adjodah, D., Allen, L., Allis, V., Baleria, G., Ballantyne, N., Van Bavel, J. J., Blunden, H., Braley, A., Bryan, C., Celniker, J., Cikara, M., & Willer, R. (2023). Megastudy identifying effective interventions to strengthen Americans'democratic attitudes. Working paper. https://doi.org/10.31219/osf.io/y79u5

Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies*, *69*, 1–13. https://doi.org/10.1016/j.electstud.2020.102199

Wagner, M. W., & Gruszczynski, M. (2018). Who gets covered? Ideological extremity and news coverage of members of the US Congress, 1993 to 2013. *Journalism & Mass Communication Quarterly*, *95*(3), 670–690. https://doi.org/10.1177/1077699017702836

Welbers, K., & Opgenhaffen, M. (2018). Presenting news on social media. *Digital Journalism*, *7*(1), 45–62. https://doi.org/10.1080/21670811.2018.1493939

Wojcieszak, M., de Leeuw, S., Menchen-Trevino, E., Lee, S., Huang-Isherwood, K. M., & Weeks, B. (2021). No polarization from partisan news: Over-time evidence from trace data. *The International Journal of Press/Politics*, *28*(3), 1–26. https://doi.org/10.1177/19401612211047194

Yang, J., Rojas, H., Wojcieszak, M., Aalberg, T., Coen, S., Curran, J., Hayashi, K., Iyengar, S., Jones, P. K., Mazzoleni, G., Papathanassopoulos, S., Rhee, J. W., Rowe, D., Soroka, S., & Tiffen, R. (2016). Why are "others" so polarized? Perceived political polarization and media use in 10 countries. *Journal of Computer-Mediated Communication*, *21*(5), 349–367. https://doi.org/10.1111/jcc4.12166

Zhao, X., Lynch Jr, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257

Zoizner, A., Shenhav, S. R., Fogel-Dror, Y., & Sheafer, T. (2021). Strategy news is good news: How journalistic coverage of politics reduces affective polarization. *Political Communication*, *38*(5), 604–623. https://doi.org/10.1080/10584609.2020.1829762

Human Communication Research, 2024, 50, 404–418
https://doi.org/10.1093/hcr/hqae006
Original Research