

The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability?

KAYLYN JACKSON SCHIFF *Purdue University, United States*

DANIEL S. SCHIFF *Purdue University, United States*

NATÁLIA S. BUENO *Emory University, United States*

This study addresses the phenomenon of misinformation about misinformation, or politicians “crying wolf” over fake news. Strategic and false claims that stories are fake news or deepfakes may benefit politicians by helping them maintain support after a scandal. We posit that this benefit, known as the “liar’s dividend,” may be achieved through two politician strategies: by invoking informational uncertainty or by encouraging oppositional rallying of core supporters. We administer five survey experiments to over 15,000 American adults detailing hypothetical politician responses to stories describing real politician scandals. We find that claims of misinformation representing both strategies raise politician support across partisan subgroups. These strategies are effective against text-based reports of scandals, but are largely ineffective against video evidence and do not reduce general trust in media. Finally, these false claims produce greater dividends for politicians than alternative responses to scandal, such as remaining silent or apologizing.


Misinformation in political discourse can negatively impact political accountability, trust, and social cohesion (Jerit and Zhao 2020; Vaccari and Chadwick 2020b). Concerns about misinformation are only deepening with the emergence of new methods to generate and disseminate falsified media, methods that are transforming and extending traditional strategies of promoting misinformation. While scholars have debated the direct effects of misinformation in terms of its ability to deceive and persuade, misinformation can serve a variety of purposes beyond direct persuasion, working through emotional and symbolic means and shifting the foundations of the broader informational environment itself. This study devotes attention to these indirect effects and provides novel experimental evidence related to one such subtle and concerning consequence of misinformation: the liar’s dividend.


In particular, we seek to understand whether politicians and other public figures can leverage an environment of misinformation and distrust to their benefit by falsely claiming that damaging true information about themselves (e.g., a scandal) is fake. That is, we explore


whether politicians can maintain support by spreading *misinformation about misinformation*—falsely claiming that true events and stories are merely “fake news” or “deepfakes.” If such lies are used successfully, they provide a benefit, or a “liar’s dividend,” increasing the liar’s authority, reelection prospects, or reputation (Chesney and Citron 2019). However, they do so through deception and risk further undermining political discourse, social cohesion, and public trust in the media and larger informational environment.

We investigate the liar’s dividend through five pre-registered¹ experimental studies, administered to a total of over 15,000 American adults. All five studies employ text-based treatments reporting four real politician scandals in the United States ($n = 11,820$), and two studies (Studies 1 and 4) additionally use video treatments of these same scandals ($n = 3,467$). We follow the politician scandals with rebuttals from the depicted politicians which state that the stories are mere misinformation—false stories or “deepfakes.” The politician claims of misinformation primarily make use of two strategies. First, politicians may seek to undermine confidence in the informational environment, invoking a channel we term “informational uncertainty.” Alternatively, they may exploit affective polarization and partisan animus to draw supporters to their defense, which we term “oppositional rallying.” We evaluate the extent to which these strategies bolster support for the politicians, impact belief in the scandal, and undermine trust in the media environment generally. We also assess whether lying pays off more for text (“fake news”) versus video (“deepfakes”) stories and

All authors contributed equally.

Corresponding author: Kaylyn Jackson Schiff , Assistant Professor, Department of Political Science, Purdue University, United States, schiffk@purdue.edu.

Daniel S. Schiff , Assistant Professor, Department of Political Science, Purdue University, United States, dschiff@purdue.edu.

Natália S. Bueno , Assistant Professor, Department of Political Science, Emory University, United States, natalia.bueno@emory.edu.

Received: April 18, 2022; revised: December 19, 2022; accepted: December 19, 2023. First published online: February 20, 2024.

¹ Our pre-analysis plans are available at <https://osf.io/qpxr8/>.

whether claims of misinformation are more effective than alternative politician responses: apologizing or simply denying (without alleging misinformation).

We find that false claims of misinformation do pay a liar's dividend, but primarily (if not exclusively) when used against scandals reported in text format. Rebuttals of scandals invoking both informational uncertainty and oppositional rallying lead to increases in politician support, and both strategies are broadly effective across partisan subgroups. Claims of misinformation also generate larger support gains for politicians than simply ignoring the scandal (non-response) or apologizing, arguably a normatively preferable strategy, and are at least as effective as a more simple denial. Additionally, we find evidence in one study that claims invoking the oppositional rallying strategy can discredit scandals caught on video. However, the effectiveness of claims of misinformation against video in one of the studies is the exception, and claims of misinformation are far more effective for text-based scandals. Meanwhile, politicians who use the informational uncertainty strategy in alleging misinformation are still likely to be held accountable when video evidence is available.

Despite the effectiveness of false claims of misinformation, especially against scandals reported in text format, these claims are not consistently accompanied by reduced belief in the scandal or reduced trust in the media. While we initially theorized that claims of misinformation would mainly operate via reduced belief in the scandal, especially for claims invoking informational uncertainty, our findings regarding belief in the scandal are mixed, and depend on the measure of belief used. Furthermore, we also originally conjectured that these claims of misinformation could further undermine trust in media, but we generally find no such impacts.

Across the five studies conducted, and based on our learning and reviewer feedback, we make a variety of modifications in substance and presentation compared to our pre-registered hypotheses and analysis plans, and our research design evolved to answer lingering questions. We encourage interested readers to review Supplementary Material (SM) Sections A.4 and A.5 for details on our pre-registered hypotheses and deviations from our pre-analysis plans, including pointers to the SM where alternative and original analyses are available.²

DIRECT AND INDIRECT HARMS OF MISINFORMATION

Fears about the impacts of misinformation are fueled by the perception of frequent, everyday encounters with misinformation: 89% of Americans report encountering made-up news at least sometimes, and

Americans are more likely to identify made-up news as a critical problem than climate change, racism, and illegal immigration (Mitchell et al. 2019). Notably, 25% of tweets spread during the 2016 U.S. presidential elections were fake or misleading (Bovet and Makse 2019), and subsequent politically oriented misinformation culminated in a violent insurrection after the 2020 election. These recent events highlight the potential for misinformation to deepen social and political fractures by exacerbating polarization, undermining accountability and rational deliberation, and decreasing trust in institutions and media as part of a vicious cycle.

While the use of misinformation for political ends is as old as politics itself (Arendt 1973; O'Shaughnessy 2004), new technological trends are potentially historically transformative. One such development is the emergence of new sophisticated methods to produce digitally altered or fabricated audio, images, or videos, known as "deepfakes," which result from advances in artificial intelligence techniques such as generative adversarial networks and diffusion models. Deepfakes are produced using approaches such as facial swapping, facial animation, and the creation of entirely synthetic images or audio. Notably, there are less sophisticated techniques referred to as "cheapfakes" or "shallowfakes" which involve basic splicing, editing, or decontextualizing of audio or visual media but still present similar risks (Barari, Lucas, and Munger 2021; Tandoc, Lim, and Ling 2018). Furthermore, advanced large language models and multi-modal models popularized in 2022, such as ChatGPT and Copilot, have significantly democratized the ability to create synthetic media. These models can generate realistic text, images, audio, and even video, making it easier for individuals to create synthetic media content that can mimic human-generated content with increasing accuracy. That is, while advanced media creation and manipulation capabilities were previously restricted to professionals through time-consuming and expensive efforts, it is increasingly straightforward for non-sophisticated actors to generate highly convincing fake video, images, audio, and text rapidly and at low cost.

The widespread accessibility of these models, including as a trend on popular social media platforms, has raised concerns about the potential misuse and manipulation of information, further highlighting the need for robust detection, verification, literacy, and governance efforts to combat the spread of misinformation and deception in digital media. Concerns surrounding deepfakes and AI-generated text-based misinformation have now permeated society: 90% of the public say altered video and images cause confusion (Mitchell et al. 2019), news media and technical experts report severe challenges in establishing the authenticity of content (Toews 2020), and politicians have raised alarm and urged legislative activity as well.

Although some experts previously suggested that the impact of deepfakes would be limited, examples of problematic uses are now proliferating in the political arena. For example, deepfakes have been used to discourage supporters of opposing parties from voting in an Indian election, to influence attitudes and military

² SM Sections, Tables, and Figures with "A" references are in the main SM file. "B" references refer to the Appendix to the SM, available with the replication materials (see Schiff, Schiff, and Bueno 2024).

responses to the Russian invasion of Ukraine, and to depict a false bombing of the U.S. Pentagon. Of particular relevance for this study, deepfakes have been used to depict specific politicians engaging in controversial acts or making offensive statements. For example, videos have allegedly exposed sex scandals involving Malaysian deputy minister Shamsul Iskander and Brazilian governor João Doria, though—critically—there has been debate in both cases regarding whether the videos are deepfakes or authentic (Toews 2020). Still other politically salient deepfakes have featured U.S. presidents Donald Trump and Barack Obama, French president Emmanuel Macron, Ukrainian president Volodymyr Zelensky, and Russian president Vladimir Putin. This content has even been shared by prominent political figures: Republican Representative Steve Scalise's office created and shared a cheapfake of U.S. president Joe Biden, Donald Trump Jr. shared a deepfake depicting his father's appearance on a CNN town hall, and the RNC used AI to generate content for an anti-Biden attack ad featuring hypothetical disaster scenarios. In among the most concerning political cases to date, an allegation that a video depicting Gabon's president as healthy was a deepfake helped to spur an unsuccessful military coup. Given these many examples, one striking implication is that the *mere existence* of deepfakes may allow for plausible claims of misinformation and lead to significant social and political harms, even when the authenticity of the content is disputed or disproved.

Notwithstanding increased attention and the seemingly consequential examples above, there is scholarly disagreement over the direct effectiveness of misinformation, conceived primarily in terms of the ability of the information to persuade. Analogous to the minimal effects hypothesis in the context of political campaigns (Kalla and Broockman 2018), some scholars have argued that the impact of fake news may be modest. According to this perspective, consumption of misinformation may be limited depending on individuals' media diets, restricted to those with strong partisan preferences, and moderated by individuals' ability to adjust for bias in news sources (Little 2018). Moreover, individual fake news messages may not be especially persuasive on their own in the face of the multitude of informational signals people receive and because fake news consumption is only a small portion of overall news and information diets (Guess, Nyhan, and Reifler 2020; Watts, Rothschild, and Mobius 2021). However, while it is possible that the direct persuasive effects of fake news may be less than feared, there is much unknown about the multiple possible direct and indirect impacts of misinformation, especially in the medium to long term (Lazer et al. 2018).

Indeed, it has long been understood that misinformation can serve a variety of purposes beyond direct persuasion about the truth of particular claims. For example, in the context of authoritarian regimes, misinformation has been used as means to signal the power of regimes or to encourage the performance of loyalty (Huang 2015; O'Shaughnessy 2004). Misinformation can also promote confusion and

skepticism. Deepfakes in particular seem especially likely to drive such indirect harms to the informational environment, as individuals may feel they are no longer able to trust their eyes and ears, engendering broader distrust in all content—whether authentic or falsified (Ternovski, Kalla, and Aronow 2022). In this sense, while deepfakes “might not always fool viewers into believing in something false,” they may exacerbate uncertainty and distrust, “further eroding our ability to meaningfully discuss public affairs” and discern truth from fiction (Vaccari and Chadwick 2020a). Appreciating that these effects may be intentional rather than merely incidental is essential if we are to understand the full implications of misinformation. As such, this article examines indirect effects of misinformation and additionally considers whether new tools to promote misinformation (deepfakes) exacerbate or otherwise alter extant challenges.

A THEORY OF THE LIAR'S DIVIDEND

In light of the importance of indirect effects, this paper is concerned with a form of misinformation that owes its existence (in part) to fake news. That is, the widespread awareness of fake news has opened the door to *false claims of fake news*, whereby politicians or other public figures can—potentially credibly—claim that real news stories are merely fake news or deepfakes, leading to what Chesney and Citron (2019) term the “liar's dividend.” While this tactic of denial and deflection has been made prominent by U.S. President Donald Trump, calls of “fake news” have now been echoed by politicians in Russia, Brazil, China, Turkey, Libya, Poland, Hungary, Thailand, Somalia, Myanmar, Syria, and Malaysia (Erlanger 2017). This form of misinformation has been used to target political opponents and to deny critical media coverage, even when objective observers and experts find the reporting to be credible. As a few notable examples amongst many, former Spanish Foreign Minister Alfonso Dastis claimed that images of police violence in Catalonia in 2017 were “fake photos” (Oppenheim 2017) and American Mayor Jim Fouts called audio tapes of him making derogatory comments toward women and Black individuals “phony, engineered tapes” (Wang 2017), despite expert confirmation.

That the systematic usage of *misinformation about misinformation*—alleging “fake news” or “deepfakes” in response to real stories—has grown in recent years suggests that public figures find this strategy to be effective or beneficial, against the expectations of a minimal effects hypothesis. In particular, politicians may believe such a strategy can be employed to avoid accountability for political abuses or scandals. We therefore hypothesize that this strategy pays off by safeguarding politicians' reputations: members of the public are less likely to penalize politicians for scandals when they “cry wolf” over fake news and deepfakes. In particular, we expect this strategy to be more effective than three alternative politician communication strategies: (1) non-response, representing an attempt to

ignore a scandal and let it blow over; (2) apologizing, arguably a normatively preferable response; or (3) simply denying a scandal without invoking misinformation.

Liar's Dividend Hypothesis: *In the face of scandal, claims of misinformation (fake news or deepfakes) will increase average support for politicians relative to no response, apologizing, or simply denying a scandal.*³

We propose that a rebuttal claiming that a story is a deepfake or fake news might improve politician support through two potential pathways, associated with respective strategies that politicians have used. First, the public may find claims of “fake news” to be credible due to uncertainty regarding the truth of signals in what many members of the public may perceive as a distorted media environment—a channel we term *informational uncertainty*.⁴ Here, the payoffs of claims of fake news result from misinformation’s truth-undermining effects, or “the principle that any information could be fake” (Ciancaglini et al. 2020). The concern is not that “people will be deceived, but that they will come to regard everything as deception” (Schwartz 2018), particularly because it is easy to challenge the veracity of evidence in a fractured political environment (Hao 2019), and harder to disprove these kinds of claims (Galston 2020). If consumers of information believe they have no credible signals about truth or falsity of political claims, the result may be increased uncertainty, as individuals lack sufficient information to establish a basic ground truth or make informed choices (Vaccari and Chadwick 2020b). Thus, even when individuals are motivated to hold accurate beliefs, informational uncertainty undermines their capacity to do so.

To illustrate how informational uncertainty operates in the case of the liar’s dividend, consider the statement of Spanish Minister Alfonso Dastis, who attempted to discredit photos of violence in Catalonia: “I’m sure you have seen what you have seen, but I have seen fake photos that date back to 2012. So, I think we have got to be patient, and look at the situation” (Oppenheim 2017). The uncertainty induced by a statement like this (perhaps intentionally) may leave citizens unclear about how to update their evaluation of the politician or scandal. More generally, after learning of an embarrassing moment or political scandal, a member of the public will be more likely to downgrade their evaluation of the politician. However, if the politician then issues a statement disclaiming the story as a deepfake or fake news, then some members of the public may be more uncertain about what is true, decreasing belief in

the scandalous story and increasing average support for the politician. We expect these effects to be concentrated among individuals in the middle of the political spectrum, representing individuals less likely to be strong supporters or opponents of partisan politicians.

The proposed channel of informational uncertainty is perhaps most active when individuals are motivated to hold true beliefs and engage in rational updating of beliefs and subsequent evaluations of politicians. In contrast, we suspect that strong supporters are especially likely to react to appeals to their partisan identity and to engage in motivated rather than purely accuracy-driven reasoning (Taber and Lodge 2006). In an environment of heightened polarization, and without credible and shared informational signals to support rational processing, individuals may be especially prone to abandoning accuracy motivations in favor of partisan “directional” ones (Druckman 2012; Pennycook et al. 2021).

These elements of the political environment are highlighted by the second channel, which we term *oppositional rallying*. To avoid cognitive dissonance in the face of identity-incongruent information (a damaging news story about a preferred politician or party), core supporters or strong co-partisans may be prone to motivated reasoning (Bullock et al. 2015). Claims that offer congenial information are useful for co-partisans who observe a scandal and seek information to justify their continued support of the politician. The allegation of a deepfake or fake news can provide just this sort of reason for supporters to rally around the politician, disregarding the negative coverage and preserving their positive evaluations of the politician. Such a response may reflect genuine changes in belief, or instead, expressive responding and partisan cheerleading (Peterson and Iyengar 2021). Further, this channel often explicitly invokes references to political opponents, making use of a “devil shift” (Sabatier, Hunter, and McLaughlin 1987) whereby politicians signal not only their own innocence, but also the guilt of political opponents and media, encouraging supporters to rally against the opposition.

As an example of this strategy, American Mayor Jim Fouts alleged that his opponents were attempting to “hijack [the annual MLK Day] ceremony by releasing more vile, vitriolic, phony tapes against me” and that such an “effort ... is designed to distract from my efforts of inclusion for all” (Wang 2017). A politician who employs the strategy of oppositional rallying may thus explicitly signal to supporters because they seek to prime partisan directional motives. We therefore expect this mechanism to be most influential when individuals have strong positive associations with a specific politician (Flynn, Nyhan, and Reifler 2017), though strong party identification alone may be sufficient to drive these effects, given increasing affective polarization and a heightened connection between partisanship and identity (West and Iyengar 2020). Thus, we expect effects to be stronger for co-partisans, who are more likely to reward rebuttals claiming misinformation that employ the oppositional rallying strategy with greater support for their preferred

³ This updated version of our primary hypothesis incorporates apologies and simple denials that were introduced after our first study was conducted, and more clearly states these reference groups in a single hypothesis for simplicity. See SM Section A.4.

⁴ We take the two politician strategies, observed in real-world behavior, to reflect distinct channels, or mechanisms, of influence. Following Gerber and Green (2012), we operationalize these two channels explicitly as separate treatments.

politician (Craig and Cossette 2020).⁵ In contrast then, strong out-partisans should be less willing to grant a dividend to politicians with an opposing political orientation, and may even prefer to punish politicians who make such claims.

Given the potential accountability concern that claims of misinformation will reduce individuals' beliefs in real politician scandals, we examine whether both informational uncertainty and oppositional rallying claims are effective in reducing beliefs about the scandal. Nonetheless, these claims of misinformation could trigger other reactions in voters that could lead to improved support for the politician. For example, through claims of misinformation, politicians might be able to signal their strength, or that they share voters' skepticism of "mainstream media."

Mediating Factors and Further Consequences

Given the expanded use and awareness of manipulated or synthetic video including deepfakes, our study also considers whether the dynamics surrounding misinformation and the liar's dividend differ for text-based versus video-based content. There is much unknown about the extent to which deepfakes constitute a major societal risk as compared to text-based fake news, and our study aims to provide helpful evidence to address that point. On the one hand, given a "psychological predisposition to believe in audio-visual content and a truth-default tendency," individuals are more likely to find video information credible (Ciancaglini et al. 2020). A "realism heuristic" implies that individuals find audio-visual content to more closely resemble real-world experience such that videos may be more naturally assimilated than text-based content (Sundar, Molina, and Cho 2021; Vaccari and Chadwick 2020b).

On the other hand, other recent studies call into question the extent to which deepfakes are more credible and persuasive than text-based misinformation (Barari, Lucas, and Munger 2021), though video-based misinformation may be more effective for changing beliefs (Wittenberg et al. 2021). Relatedly, there is a long-standing debate about the extent to which "vivid" content—referring to the ability of information to provoke emotions and interest—is actually more persuasive (Taylor and Thompson 1982). In the context of the liar's dividend, on balance, we hypothesize that respondents will believe that videos depicting politician scandals are harder to fake than text, such that claims that these videos are deepfakes will be perceived as less credible, translating into a smaller payoff for politicians.

Deepfakes Hypothesis: *Rebuttals claiming misinformation, relative to no response, will lead to smaller improvements in average support for politicians when the underlying stories are reported via video as compared to text.*

⁵ Claims of misinformation may be less effective for clientelistic co-partisans whose support is conditional upon expected benefits and more disconnected from evaluations of the politician.

An environment of pervasive distrust in media and institutions, itself partly fostered by misinformation, constitutes a fertile ground for the liar's dividend. Indeed, the liar's dividend takes advantage of confusion regarding source credibility, making it more plausible that politician claims of misinformation are perceived as credible (Desai, Pilditch, and Madsen 2020). If claims of misinformation further reduce trust in media, then they can create the conditions for avoiding accountability not only for today's scandal, but for tomorrow's as well. Consistent with previous studies (Ladd 2011), we hypothesize that elite, partisan claims of misinformation may decrease general trust in media.

Trust in Media Hypothesis: *Rebuttals claiming misinformation will lead to decreased trust in media relative to no response, denial, and apologizing.*⁶

EXPERIMENTAL DESIGN

To address the hypotheses presented above, we conducted five IRB-approved, online survey experiments based on pre-registered designs⁷ with a total of 15,287 respondents.⁸ The studies consider how Americans react to politicians' claims of misinformation in response to scandalous news stories presented via text (in Studies 1–5) or video (in Studies 1 and 4). We randomly assigned participants, irrespective of political party, to a real scandal⁹ involving one of four politicians—two Democrat and two Republican—making statements that are arguably insensitive or embarrassing.¹⁰

In Studies 1, 2, 4, and 5, after viewing the politician scandal, participants were then also randomly assigned to one of three politician responses: no response (control), a rebuttal claiming misinformation which primes informational uncertainty (IU), or a rebuttal claiming misinformation which primes oppositional rallying (OR).¹¹ Thus, in Studies 1, 2, 4, and 5, the control non-response represents a politician strategy of ignoring the scandal in hopes that it will blow over. In Study 3, we

⁶ This hypothesis is also updated to incorporate expectations across our multiple studies.

⁷ This article incorporates some language previously included in our pre-analysis plans.

⁸ We chose the sample size for the first study based on minimum detectable effect calculations using results from a pilot study in August 2020 (see SM Section B.6). For subsequent studies, we performed power analyses using results from the prior studies.

⁹ Of note, the scandals studied here relate to identity politics, which has some bearing on the generalizability of our findings. See SM Section B.13 for a discussion.

¹⁰ We use stories of former politicians to ensure minimal impacts on current officials. Based on pilot results discussed in SM Section B.13, we identified four stories that respondents viewed as similarly embarrassing and plausibly digitally faked. We also selected clips that were as consistent as possible given available options in terms of length, content, and context.

¹¹ While recent discourse surrounding "fake news" provides the context for this study, this term is polarizing, as we confirmed in a pilot study (see SM Section B.13). We instead use "false and misleading" in our treatments, a phrase also commonly used by politicians when claiming misinformation.

TABLE 1. Summary Information about Survey Experiments

Study	Date	Platform	<i>n</i>	Scandal format	Politician response treatments
Study 1	Feb. 2021	Lucid	2,503	Text or video	Nonresponse, IU, OR
Study 2	May 2021	Lucid	2,518	Text only	Nonresponse, IU; nonresponse, OR
Study 3	Oct. 2021	Lucid	2,996	Text only	IU, apology, simple denial
Study 4	Nov. 2022	Prolific	4,432	Text or video	Nonresponse, IU, OR
Study 5	Nov. 2022	Lucid	2,838	Text only	Nonresponse, IU, OR

instead consider how claims of misinformation invoking informational uncertainty compare to two other politician response strategies: an apology and a simple denial without an allegation of misinformation. We include the simple denial and apology treatments in Study 3 to assess whether the current informational environment makes claims of misinformation even more effective than alternative longstanding politician responses to scandal. Table 1 provides an overview of the five studies.

For the video, text, and politician rebuttal treatments that we describe next, we aimed to reduce media source cues (e.g., by cropping news banners from videos) to (1) maintain symmetry across treatments and improve internal validity and (2) focus on respondent identification with the politician and party rather than the particular media source. While we recognize that such a strategy removes an element of realism in normal news consumption, we thought this approach best balanced concerns about internal versus external validity. Reassuringly, experimental designs with sparser details, albeit less naturalistic, tend to enable researchers to identify the existence of an effect and do not necessarily imply less generalizability (Brutger et al. 2023). Further, the use of multiple stories and averaging results across politicians helps to ensure that our findings are not limited to a single media source, politician, scandal, or political party.¹²

The rebuttals alleging misinformation are inspired by real politician statements and are designed to invoke considerations related to informational uncertainty and oppositional rallying, but are not strictly derived from statements made by the depicted politicians themselves. The informational uncertainty rebuttal draws from comments such as those made by Foreign Minister Dastis and by Syrian President Bashar al-Assad, who in an attempt to discredit an Amnesty International report said: “You can forge anything these days... We are living in a fake news era” (Erlanger 2017). Along these lines, participants in the informational uncertainty treatment group saw the claim in Figure 1a.

For the oppositional rallying rebuttal, we drew inspiration from statements like that of Mayor Fouts, along with comments by then-president Donald Trump on Twitter in response to growing criticism over his handling of the pandemic: “The Fake News Media and their partner, the Democrat Party, is doing everything within its semi-considerable power (it used to be

¹² Our treatment effect estimates reflect averages across politicians. Results with politician fixed effects and heterogeneous treatment effects by politician are very similar, as shown in SM Section B.12.

FIGURE 1. Treatment Wordings for Claims of Misinformation

[Politician Name] Responds That Story is False and Misleading, People Should Be Skeptical

In response to the recent allegations, [Republican | Democrat] [Politician Name] asserted that the story is false and misleading. He claimed that [the video is a deepfake, a computer-edited video that uses fake audio and images | the story is not based on true information].

When asked about the incident, he said that it's well known that there's a lot of misleading information, so people should be skeptical about what they hear.

[Last Name] stated that “You can't know what's true these days with so much misinformation out there.”

(a) Informational Uncertainty Treatment

[Politician Name] Responds That Story is False and Misleading, Attack by Opponent

In response to the recent allegations, [Republican | Democrat] [Politician Name] asserted that the story is false and misleading. He claimed that [the video is a deepfake, a computer-edited video that uses fake audio and images | the story is not based on true information].

When asked about the incident, he said that the story is an attack by the opposition, and that people should not pay attention to it.

[Last Name] stated that “My opponent would say anything to hurt me, but my supporters know who's really on their side.”

(b) Oppositional Rallying Treatment

greater!) to inflame the CoronaVirus situation.” Participants in the oppositional rallying treatment group saw the claim in Figure 1b.

An important aspect of our study is that it addresses sensitive social and political issues in the context of misinformation, an already fraught topic in that interaction with misinformation can harm participants. As such, we carefully considered ethics in the design and administration of our surveys. Foremost, our study unavoidably involved deception, given our focus on misinformation and the liar's dividend. Our approach to minimizing deception was to use real videos and stories of politicians, rather than, for example, generating a new deepfake or false story. To enable the comparison of different

politician communication strategies in the context of misinformation, the research team did attribute various responses to the politicians (e.g., a rebuttal claiming misinformation or apology) that they did not actually make. Given this deception, we debriefed all participants at the end of the study. Second, to avoid exacerbating participant feelings of distrust and uncertainty, our debrief included links to resources on media literacy and digital literacy, such as information on how to spot false news stories.¹³ Third, we wanted to avoid the risk that participation in the study would influence real-world political behavior such as voting. Therefore, we chose to use stories about inactive politicians, that is, individuals who are not currently in office or running for office. Fourth, all participants consented prior to participation and were compensated through our survey vendors. Participants were warned that some of the information was offensive, that some information would be withheld, and that additional information about the goals of the study would be provided at the end.

We measure respondents' belief in the underlying scandal using two outcome questions ("I believe the story about the politician" and "I think that the story about the politician is true").¹⁴ We then use a set of four outcome measures to assess whether respondents supported the politician ("I would support the politician," "I would defend the politician against critics," "I would vote for the politician," and "I would donate to the politician"). Finally, we assess respondents' trust in media using two outcome questions ("I trust the media" and "I believe that the media reports the news fairly"). All outcome questions use a bipolar 5-point Likert scale with respondents indicating their agreement from "Strongly disagree" to "Strongly agree." With the goal of reducing variance and improving content validity, we use multiple questions and create pre-registered indices for each outcome.¹⁵

To test our hypotheses, we regress the corresponding outcome measure (e.g., the politician support index) on treatment, along with a set of covariates (partisanship, gender, race/ethnicity, age, education, household income, region, media literacy, and digital literacy) to improve precision.¹⁶ Our hypotheses and regression specifications are pre-registered and are available at

<https://osf.io/qpxr8/>.¹⁷ For the primary regressions used to test our hypotheses, we report standard two-sided *p*-values based on robust standard errors.¹⁸

The samples for Studies 1–3 and 5 were recruited using the Lucid Theorem platform and are demographically proportionate to the U.S. adult population in terms of gender, race/ethnicity, age, and region. Study 4 uses the Prolific platform and its sample is proportionate to the U.S. adult population on gender, race/ethnicity, and age. See SM Table A1 for information about the composition of our samples. We also find no evidence of covariate imbalance in our samples.¹⁹ We include two attention screener questions to allow for the analysis of results stratified by level of attentiveness of respondents, shown in SM Section B.4 (Berinsky et al. 2019). To avoid selection bias, our main analyses do not exclude inattentive respondents, as recommended by Berinsky, Margolis, and Sances (2014). Instead, we report results for the full samples as well as for the attentive subsets. Overall, our findings are consistent, if not stronger, among more attentive participants (also see SM Sections A.2 and A.6).

IS THERE A LIAR'S DIVIDEND?

We conducted Study 1 in February 2021 with 2,503 respondents, and we also conducted an exact replication of Study 1, which we label Study 4, in November 2022 with 4,432 respondents. Figure 2 presents the 2×3 factorial design used in both studies, with variation in both the presentation of the politician scandal (text or video²⁰) and the subsequent politician response (no response, a rebuttal invoking informational uncertainty, or a rebuttal invoking oppositional rallying). In order to ensure consistency across the text and video treatments, we create transcripts of the video clips to produce the text-based treatments. The design of Studies 1 and 4 thus allows for examination of the Liar's Dividend Hypothesis, as well as the Deepfakes Hypothesis, through exploring differences in responses to politician claims of misinformation after video-based versus text-based scandals.

Figure 3 presents standardized treatment effects for Studies 1 and 4 in order to assess the impact of claims of misinformation on politician support. Overall, the results provide strong support for the Liar's Dividend

¹³ Information about our survey instruments is available in SM Sections B.1–B.4.

¹⁴ We introduce an additional belief measure in Study 5 that asks participants about whether they believe that the politician made the specific remarks in the scandal.

¹⁵ The indices are constructed following a pre-registered procedure used by Kling, Liebman, and Katz (2007) which involves averaging *z*-scores for the component outcome questions.

¹⁶ We discuss the news media literacy and digital literacy measures in SM Section B.4. We code more levels of race/ethnicity than originally pre-registered in order to identify distinct subgroups of respondents and report this deviation in SM Table A3. Results with the original coding are available in our replication materials. Covariate-unadjusted main results are included in SM Section B.12. We do not weight the survey respondents in our samples, as Lucid and Prolific provide samples that are already balanced to be similar to the U.S. population on several—but not all—important observable characteristics (see SM Table A1 for more details).

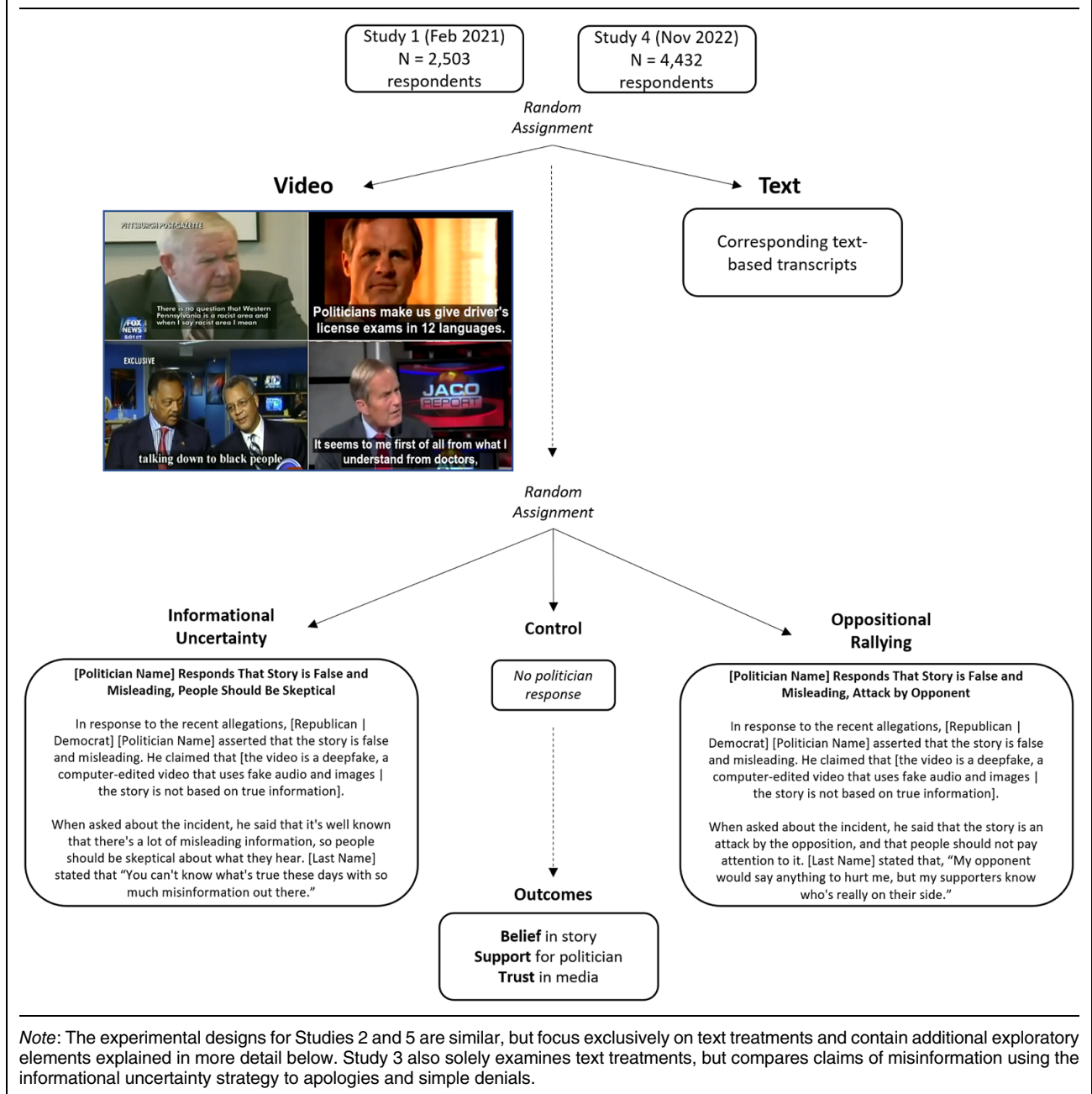
¹⁷ We originally pre-registered analyses pooling the text and video treatments, but present these separately for clarity based on reviewer suggestions. Similarly, we also originally pre-registered pooling the informational uncertainty and oppositional rallying strategies into a single "allegation" treatment due to power concerns, but again report separate results for clarity based on reviewer feedback. See SM Section A.5 for more details. SM Section A.6 also includes results based on both alternative approaches to pooling.

¹⁸ We also engage in exploratory analyses using corrections for multiple testing (see SM Section A.6).

¹⁹ See SM Section B.5 for covariate balance information.

²⁰ For all studies, we used a timer to require individuals to engage for at least 10 seconds (about the length of the videos) before advancing. We also added subtitles to the videos embedded in the survey. Click tracking in Study 4 showed that 99% of individuals clicked to watch the embedded videos.

FIGURE 2. Experimental Design for Studies 1 and 4



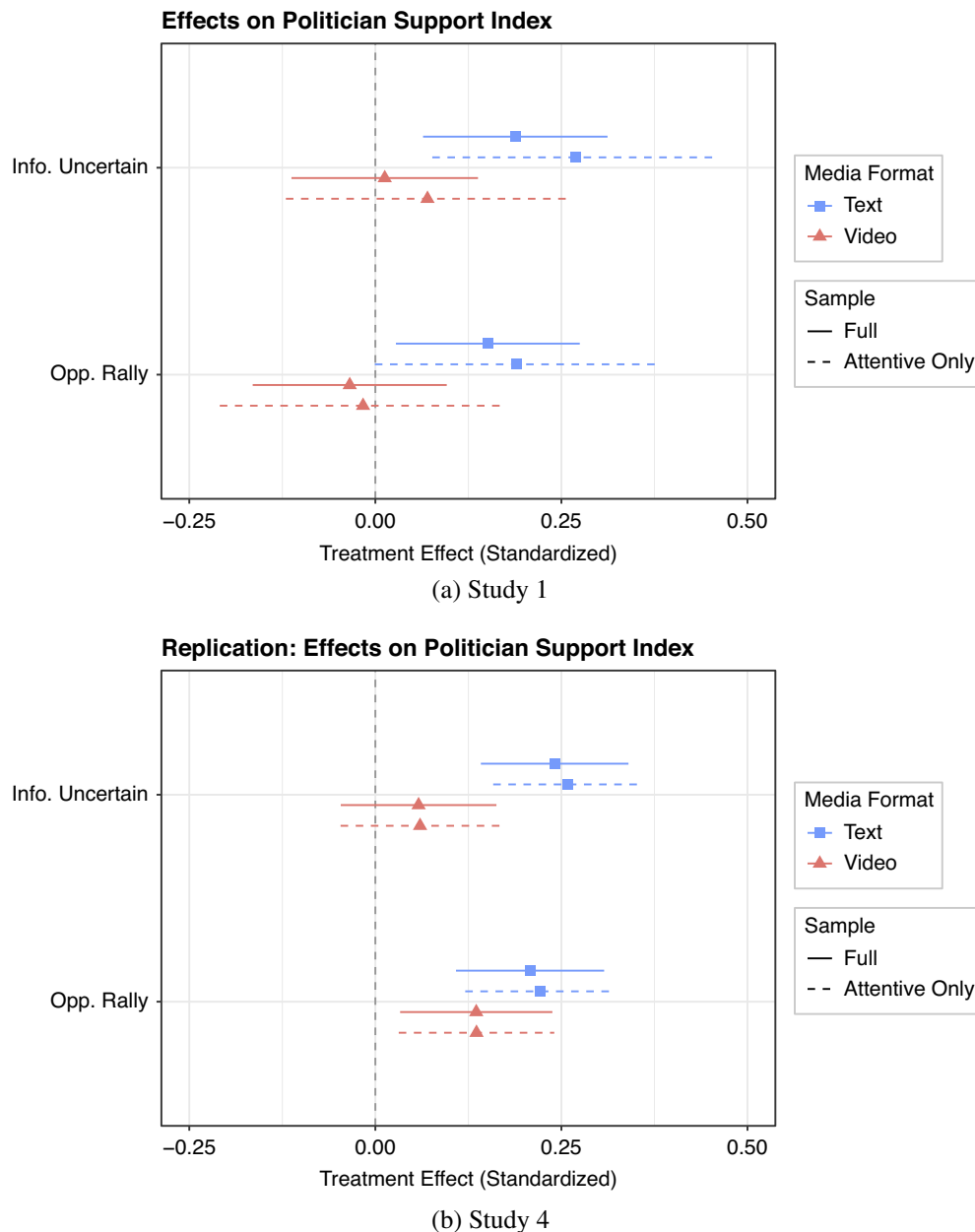
Hypothesis, at least for text-based scandals. Figure 3a and 3b show that claims of misinformation invoking informational uncertainty significantly increase politician support by around 0.19 (Study 1) to 0.24 (Study 4) standard deviations after text-based scandals, but fail to increase politician support after scandals caught on video. Claims invoking oppositional rallying also significantly increase politician support by around 0.15 (Study 1) to 0.21 (Study 4) standard deviations after text-based scandals. Results are mixed, however, for oppositional rallying claims after video scandals, with insignificant effects in Study 1 but significant increases

in politician support (by 0.14 standard deviations) in Study 4.²¹

The magnitudes of these liar's dividend effects for text-based scandals are meaningful, with effect sizes of 0.1 considered "small" and 0.2 "medium" in the political psychology literature (Funder and Ozer 2019). For context, the largest treatment effect for a single component outcome measure—for example, "I

²¹ These effects that we report in the text are for the full sample, but Figure 3 also shows effects for the attentive subset.

FIGURE 3. Liar's Dividend Results for Studies 1 and 4



Note: Based on the Study 1 (Lucid, Feb. 2021, $n = 2,503$) and Study 4 (Prolific, Nov. 2022, $n = 4,432$) samples. All figures display 95% confidence intervals based on robust standard errors. The reference group is composed of respondents who received a non-response from the politician. Full tables of results with covariates available as SM Tables B6 (Study 1), B7 (Study 1, Attentive), B8 (Study 4), and B9 (Study 4, Attentive).

would support the politician”—corresponds to an *unstandardized* 0.26 point (Study 1) and 0.29 point (Study 4) increase in support along the 5-point Likert scale, for text-based scandals. Another way of making sense of these effects is to examine the impact of the claims of misinformation on opponents of the politicians. In the text control group in Study 1, around 44% of respondents were opponents, measured as the percentage of respondents who disagreed or strongly disagreed that they would “support the politician.” In

contrast, claims of misinformation substantially decreased the percentage of opponents in the text treatment groups to around 32%–34%, a 10–12 percentage point reduction.²²

²² We found similar results in Study 4. 58% of respondents in the text control group in Study 4 were opponents compared to around 49% in both treatment groups (9 percentage-point reduction).

Therefore, we find substantial and significant evidence that both strategies for claiming misinformation—invoking informational uncertainty or oppositional rallying—produce a liar’s dividend after text-based scandals. Notably, effects are even larger for three of the four support measures used to create the index—willingness to support, defend, and vote for the politician—while reticence to donate to the politician attenuates combined support as measured by the index (see SM Figure B7). Moreover, Figure 3 reveals that estimates do not vary dramatically by survey-taker attentiveness and are consistent across Studies 1 and 4.

Video and Text Scandals

While we find that politicians can effectively make claims of misinformation in response to text-based scandals, regardless of strategy, we find that claims of misinformation in response to *video* scandals are ineffective in most cases. In Study 1, we find that claims of deepfakes, using either strategy, do not increase support for the politician when the scandal is depicted in video. The treatment effects on politician support are substantially larger for text-based versus video scandals, and these differences are statistically significant for both informational uncertainty ($p = 0.048$) and oppositional rallying ($p = 0.041$). The smaller effects of claims of misinformation against video-based scandals are consistent with the Deepfakes Hypothesis, though we did not hypothesize that claims of deepfakes would fail to engender support at all.

However, we find different results in Study 4. The evidence from Study 4 suggests that claims of misinformation priming oppositional rallying are effective against video scandals, while claims invoking informational uncertainty are not. That is, the Study 4 results imply that the payoffs for claims of deepfakes might be dependent on the misinformation strategy. There are several possible explanations for the difference in effectiveness of rebuttals invoking oppositional rallying against video-based scandals between Studies 1 and 4. One possibility is increased familiarity with deepfakes, as Studies 1 and 4 were almost 2 years apart. We do find some evidence of increased awareness of deepfakes—22% of respondents in Study 1 reported that they had not heard about algorithms to produce fake video content (deepfakes), while only 8% of respondents in Study 4 had not heard about deepfakes.

Another possibility concerns the timing and context of Study 4, which occurred just after the 2022 elections. We theorized that oppositional rallying works due to partisan motivated reasoning, and heightened partisan identification and polarization around the elections (Singh and Thornton 2019) could have created a more propitious context for the oppositional rallying strategy, even against potentially more credible video evidence. We again find some evidence for this possibility—the proportion of self-reported independents is significantly lower in Study 4 (23%) compared to Study 1 (27%) ($p < 0.001$), though the difference is small. Finally, there may simply be differences in the types of respondents in the two samples, affecting

partisanship, awareness of deepfakes, or even engagement with the video treatments.²³

Although we cannot definitively determine the source of the greater liar’s dividend against video for oppositional rallying in Study 4 versus Study 1, overall, our two studies find that claims of misinformation are less effective against video-based versus text-based scandals. We theorized that claims of deepfakes are less credible because individuals are more likely to believe video content. However, we also note that part of the difference between text-based and video-based scandals may also result from an asymmetry in the treatment wordings. For the video treatment, the politician actively alleges that the video is a deepfake, while in the text treatment, the politician denies the truth of the story while claiming misinformation. It is possible that a negative statement (denial claiming misinformation) is perceived as more plausible than a positive statement (deepfake), particularly as the latter statement implies someone would have had to actively create a deepfake, rather than merely perpetuate textual context that was not accurate.

The reduced (or non-existent) effectiveness of claims of misinformation against video-based scandals is somewhat reassuring. While scholars and the public are justifiably concerned about misinformation perpetuated through the use of ultra-realistic deepfakes, an interesting irony is that video content may be so believable that politicians gain little ground when trying to pretend that authentic video content is faked. Yet, to the extent that public figures find themselves increasingly needing to rebut actual deepfakes, they may find there is no truth-teller’s dividend either.

WHY IS THERE A LIAR’S DIVIDEND?

In Studies 2 and 5, we replicate key elements of Studies 1 and 4 and also probe mechanisms behind the liar’s dividend. Studies 2 and 5 differ from Studies 1 and 4 by focusing on text exclusively (no video treatment), as prior findings showed weaker or non-existent liar’s dividend effects for video-based scandals.

For Study 2, we recruited 2,518 participants in April 2021 via Lucid. Participants were randomly assigned to one of the four text-based politician scandals, followed by: no response from the politician (control) or the informational uncertainty rebuttal.²⁴ After seeing one of the two responses to the politician scandal, participants answered the same outcome questions as in Study 1 to preserve the integrity of the replication for the informational uncertainty treatment. We followed this component of Study 2 with a secondary experiment

²³ While we obtain similar results for the text treatment in Studies 1 and 4, the higher-quality survey-takers through Prolific in Study 4 may have engaged more with the video treatments due to experience watching videos through surveys or other unobservable factors. 99% of the respondents in Study 4 clicked to watch the video treatments. This was not measured in Study 1.

²⁴ We also included a fact-checking treatment in Study 2, which we return to in the Conclusion and SM Section A.6.

embedded in the same survey to separately assess the oppositional rallying treatment. In particular, Study 2 participants were also assigned to a *second* politician scandal followed by either the control condition or oppositional rallying rebuttal, and accompanied by the support outcome questions.²⁵

The objective for Study 5 was to better understand whether individuals' beliefs about the scandal were impacted by the claims of misinformation. We recruited 2,838 participants in November 2022 via Lucid. We modified the survey questions for Studies 1 and 4 by including a more specific, clear question about belief in the scandal, described in greater detail below, and by directly asking respondents whether they believed the politician's claim of misinformation.

In the subsections below, we first present results across studies showing strong, consistent liar's dividend effects in the case of text-based scandals. Then, we explore the mechanisms associated with the oppositional rallying and informational uncertainty strategies by focusing on heterogeneous effects by co-partisanship with the politician and belief in the scandal, respectively. Finally, we discuss impacts on trust in media.

Liar's Dividend Effects across Studies

Figure 4 shows results for Studies 1, 2, 4, and 5. Figure 4 also presents "pooled" estimates, combining evidence across studies, that are precision-weighted averages of the treatment effects from each study using fixed effects specifications.²⁶

Across study arms with text-based treatments, there is strong support for the Liar's Dividend Hypothesis. As depicted in Figure 4, in seven out of eight cases across four studies, participants who were exposed to claims of misinformation reported higher average levels of willingness to support the politicians when the claims followed a scandal reported in text format. Figure 4a reveals highly consistent liar's dividend effects for text scandals, with pooled treatment effect estimates of 0.17 and 0.21 standard deviations for the informational uncertainty and oppositional rallying strategies, respectively, and with consistent, if not stronger, results for the attentive subsets of respondents. Overall, given that these dividends are produced through a single politician rebuttal, the gains in politician support are substantial.

Further, as discussed above, Figure 4b shows a more recent liar's dividend effect for video scandals in the case of the oppositional rallying strategy in Study 4. However, this effect for video represents only one non-null effect out of four cases across two studies. On balance, there is limited evidence that claims of

deepfakes in the face of audio-visual evidence boost politician support.

Co-Partisanship with the Politician

Moreover, we find that claims of misinformation are broadly effective against text-based scandals, producing support gains from not only co-partisans, but also independents and even out-partisans. In the case of oppositional rallying, we hypothesized that claims of misinformation invoking friends and foes would prime partisan political identity and stir up negative sentiments toward perceived political opponents. As a result, we expected rebuttals invoking oppositional rallying to produce the strongest effects for sympathetic co-partisans of the politician. Alternatively, in the case of informational uncertainty, we expected rebuttals invoking this strategy to produce stronger effects on independents, whose lack of partisan attachments may make them more susceptible to feelings of uncertainty and more willing to change their support.

Yet Figure 5 reveals more extensive effects. Figure 5 displays heterogeneous effects of both theoretical channels of the liar's dividend compared to control. Effects are disaggregated by the co-partisanship of respondents with the politician in their respective treatment, and are produced by pooling respondents across studies and focusing on text scandals for comparability.²⁷

In line with our expectations, we find that politician claims of misinformation designed to shift the focus to opponents result in significant support gains amongst co-partisans. That is, oppositional rallying produces sizable effects on the order of 0.28 standard deviations for co-partisans. However, these effects are not statistically different from the effects for independents, and while they are statistically different from the effects for out-partisans ($p = 0.028$), the difference is insignificant in the attentive subset ($p = 0.159$). Notably, even the gains for out-partisans are relatively substantial and statistically significant (0.17 standard deviations, $p < 0.001$), suggesting that oppositional rallying has even more widespread appeal than anticipated with no evident backlash effect for out-partisans.

Meanwhile, Figure 5 also suggests that, against expectations, independents are not impacted to a greater extent by informational uncertainty than co-partisans. However, rebuttals invoking informational uncertainty do appear relatively less effective for out-partisans, although this difference is much less pronounced and statistically insignificant in the attentive subset.²⁸ The informational uncertainty strategy may operate across the partisan spectrum because an environment of heightened uncertainty and distrust of media leads individuals to perceive politicians as

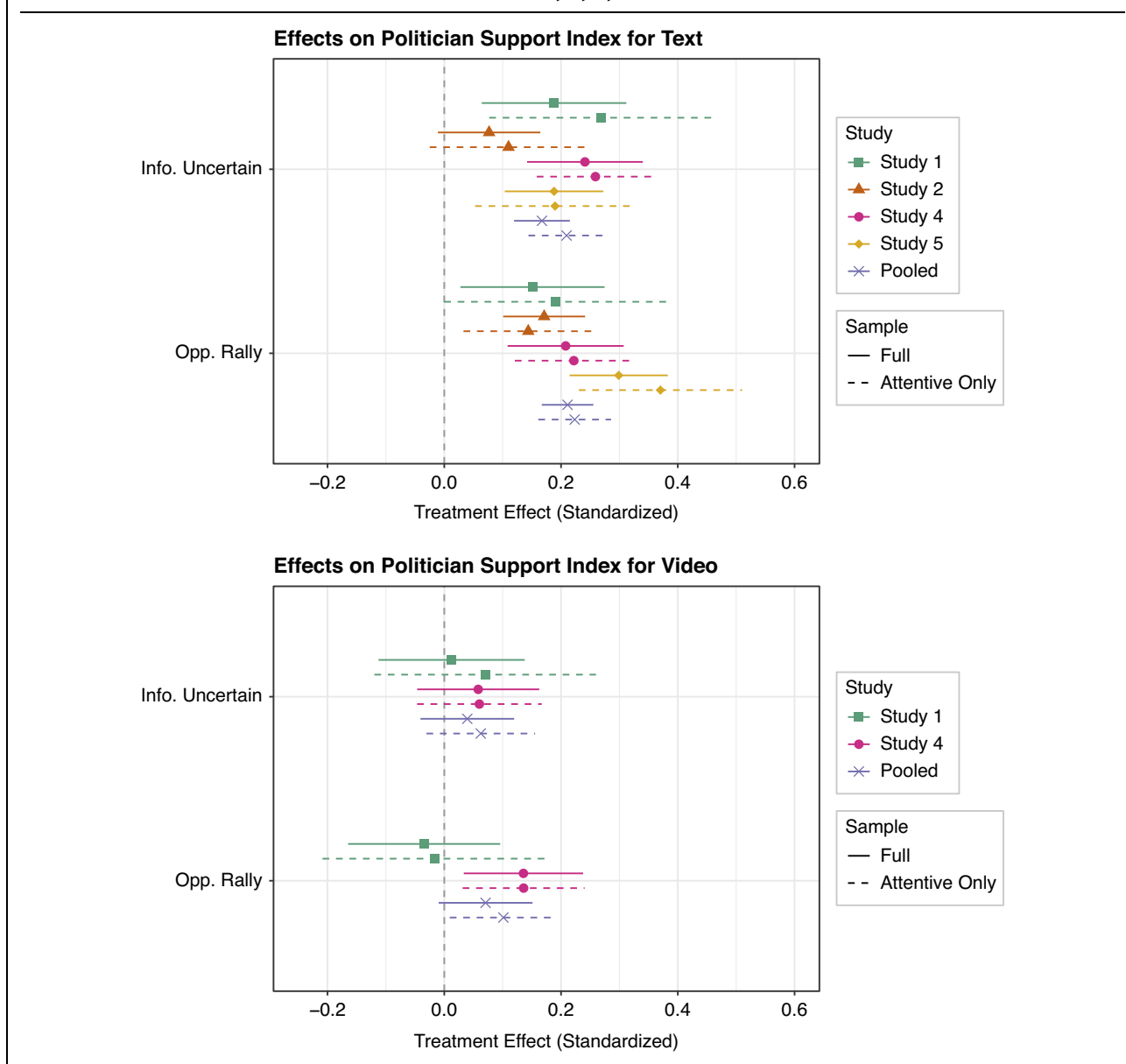
²⁵ We additionally modified the wording of the oppositional rallying treatment to avoid raising survey-taker suspicion that the study was manufactured by researchers.

²⁶ Results are consistent with random effects and with a single regression pooling samples.

²⁷ SM Section A.3 also presents results separately for each study.

²⁸ We use an alternative coding of co-partisanship than originally pre-registered in order to better identify distinct subgroups of respondents, such as out-partisans. SM Table A3 provides more detail. Also, Figure A8 presents results for alternative specifications of co-partisanship and partisanship.

FIGURE 4. Liar’s Dividend Results for Studies 1, 2, 4, and 5



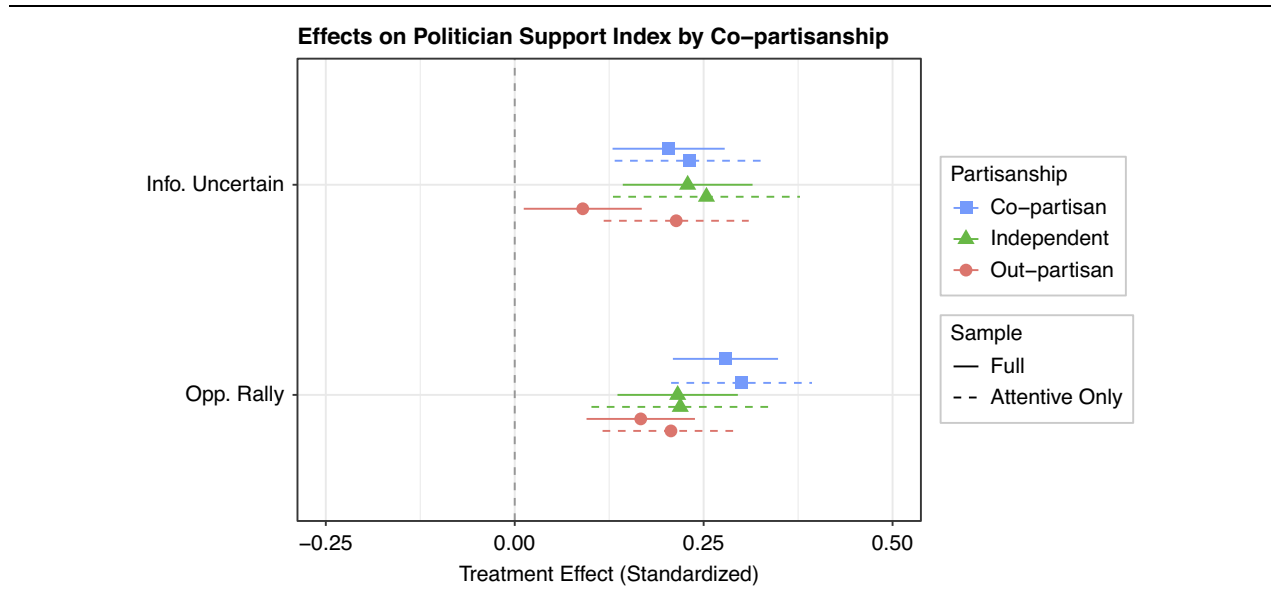
Note: Based on the Study 1 (Lucid, Feb. 2021, $n = 1,249$ Text, $n = 1,254$ Video), Study 2 (Lucid, May 2021, $n = 2,518$ Text), Study 4 (Prolific, Nov. 2022, $n = 2,151$ Text, $n = 2,146$ Video), and Study 5 (Lucid, Nov. 2022, $n = 2,838$ Text) samples. The pooled estimates are precision-weighted averages of treatment effects from each study using fixed effects specifications. We show results separately for text and video. Full tables of results with covariates available as SM Tables B10 (Text), B11 (Text, Attentive), B12 (Video), and B13 (Video, Attentive).

relatively more reliable sources than media organizations (Desai, Pilditch, and Madsen 2020).

While the effects of claims of misinformation are widespread across partisan groups, baseline levels of support vary substantially. Consequently, claims of misinformation appear to move out-partisans, on average, from the “strongly disagree” to the “disagree” category for politician support. In contrast, claims of misinformation appear to move co-partisans, on

average, from a neutral position to support (see SM Figures A4–A6). Critically, we also find consistent effects for both strategies when considering Republican and Democratic respondents separately (see SM Figure A8).

In sum, claims of misinformation against text-based scandals are broadly effective, producing liar’s dividend effects that are not solely concentrated within groups determined by co-partisanship or partisanship.

FIGURE 5. Heterogeneous Effects of Oppositional Rallying and Informational Uncertainty in Response to Text-Based Scandals

Note: Based on pooling the Study 1 (Lucid, Feb. 2021, $n = 1,249$ Text), Study 2 (Lucid, May 2021, $n = 2,518$ Text), Study 4 (Prolific, Nov. 2022, $n = 2,151$ Text), and Study 5 (Lucid, Nov. 2022, $n = 2,838$ Text) samples. Co-partisans are respondents whose self-reported partisanship matches that of the politician whose scandal they read. Out-partisan respondents are from the opposing political party to treatment politicians. For example, self-identified Strong Democrats, Democrats, and Lean Democrats are identified as co-partisans with the Democrat politicians and out-partisans with the Republican politicians depicted in the treatments. Those who identify as independents are classified as independents regardless of the politician party. This coding is a deviation from our pre-analysis plan, reported in SM Table A3. Full tables of results with covariates available as SM Tables B14 and B15 (Attentive sample). SM Section A.3 presents results separately for each study and also explores robustness to alternative specifications of co-partisanship and partisanship.

While surprising, these findings are consistent with recent findings showing that “persuasive information” (whether true or not) moves individuals who belong to different groups in the same direction and to a similar extent (Coppock 2022; Fowler and Howell 2023). Even if our treatments, especially oppositional rallying, contain partisan group cues, recent research also suggests that group cues do not necessarily undermine the effects of persuasive messages (Tappin, Berinsky, and Rand 2023).

Belief in the Scandal

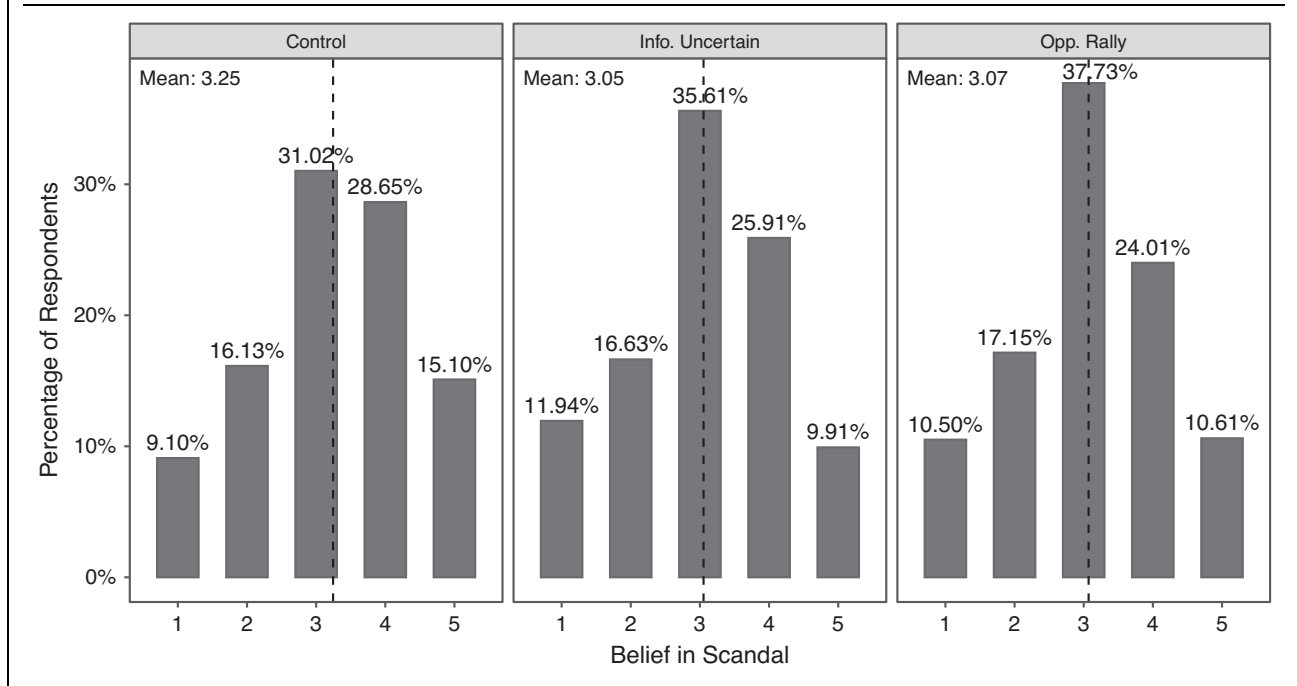
We also expected claims of misinformation invoking informational uncertainty to influence belief in the underlying scandal about the politician, lowering belief in the scandal and resulting in higher politician support. While Studies 1 and 2 (see SM Section B.12) and Study 4 showed insignificant impacts on belief in the scandal, we employed an alternative and clearer measure in Study 5 to better assess impacts on belief in the scandal. The Study 5 belief measure asks respondents whether they agree with the statement that the “politician really said” the arguably offensive comment in the scandal, using the actual wording from the scandal, thereby reducing ambiguity in our prior measure regarding whether belief refers to belief in the underlying scandal or in the claim of

misinformation.²⁹ Figure 6 presents the distributions of this new belief measure by treatment group in Study 5, which included only text treatments.

Figure 6 reveals that the politicians’ claims of misinformation did reduce belief in the underlying scandal in Study 5. While 44% of individuals in control believed the scandal (scores of 4 or 5 on the outcome scale), only 36% and 35% of individuals who received the claims of misinformation invoking informational uncertainty and oppositional rallying, respectively, believed that the politicians actually made the scandalous remarks. Figure 6 also shows a higher percentage of individuals responding with uncertainty (score of 3) in the treatment groups. These impacts are statistically significant and relatively substantial, as informational uncertainty reduced belief by 0.15 standard deviations ($p = 0.001$), and oppositional rallying reduced belief by 0.14 standard deviations ($p = 0.002$). Moreover, we find that believing the politician’s claim is strongly and negatively correlated with belief in the scandal ($r = -0.22$, and $r = -0.40$ for the attentive subset). That is, individuals in treatment who reported greater belief in the

²⁹ We pre-registered a pre-post belief change measure for Study 5 but prefer this single post belief measure because we found it to be less ambiguous, more valid, and clearer for presentation purposes. SM Section A.5 provides more detail on this deviation, and SM Table A14 reports results using the belief change measure.

FIGURE 6. Impacts on Belief in the Scandal



Note: Based on the Study 5 (Lucid, Nov. 2022, $n = 2,838$) sample, which included only text treatments. Belief in the scandal is measured by asking respondents whether they agree with the statement that the “politician really said” the arguably offensive comment in the scandal (using the actual wording from the scandal), and ranges from Strongly disagree (1) to Strongly agree (5). SM Figure B1 presents results for the attentive subset of respondents and shows even stronger effects.

politician’s claim of misinformation were also less likely to believe that the scandal had really occurred.³⁰ Finally, we also find that believing the rebuttal is strongly and positively associated with support for the politician ($r = 0.39$, and $r = 0.43$ for the attentive subset). While this could be driven by reverse correlation, it echoes a finding from Study 2 that individuals who believed the politician rebuttals were also much more likely to report that the rebuttal affected their support for the politician (see SM Table A9). Thus, in Study 5, we find evidence for the theorized pathway for informational uncertainty: (1) that claims invoking informational uncertainty reduce belief in the scandal, (2) that belief in the claim of misinformation is associated with the reduction in belief in the scandal, and (3) that belief in the politician claim of misinformation is also associated with increased politician support. While we expected that change in belief would be associated with the informational uncertainty strategy, evidence from Study 5 also suggests that the oppositional rallying strategy reduced belief in the scandal as well.

While findings from Study 5 offer clear support for the influence of rebuttals invoking misinformation on belief in the scandal, the largely mixed and null results

regarding belief in Studies 1–4 urge caution. The mixed results could stem from ambiguity in question wording or could be due to weakly held beliefs in survey-measured attitudes generally (Graham 2021). Given our mixed and null results across our original versus our newer and preferred measures of belief, we emphasize that the findings about belief depend on the measure used, and that more research is needed to evaluate this pathway. Furthermore, there are likely additional pathways through which claims of misinformation influence politician support independently of changes in belief in the scandal. For example, claims of misinformation could change individuals’ views of the politicians’ motivation to stay in office, or the politician’s ability and willingness to challenge conventional media. Consequently, changes in views about the politician could then be positively associated with politician support.³¹

Trust in Media

Beyond the immediate dividends to politicians, does misinformation about misinformation produce additional, more indirect consequences for society as a

³⁰ To entertain a competing explanation, we do not find as strong of a correlation between belief in the scandal and respondents’ impressions that the scandal is damaging to the politician’s reputation ($r = 0.14$, and $r = 0.06$ for the attentive subset).

³¹ Furthermore, we do not conduct a formal mediation analysis in which belief is a mechanism between treatment and politician support, as we find the assumptions not sufficiently credible in this case. Consequently, we measure belief solely as an outcome and interpret it as a plausible, yet not unique, pathway that influences politician support.

TABLE 2. Impacts on Trust in Media

	Trust in Media Index			
	(1)	(2)	(3)	(4)
Info. Uncertain	−0.046 (0.062)	−0.076 (0.042)	0.018 (0.048)	−0.028 (0.041)
Opp. Rally	−0.045 (0.062)		0.029 (0.048)	−0.035 (0.042)
Sample	Study 1 text	Study 2	Study 4 text	Study 5
<i>N</i>	1,249	2,518	2,151	2,838
<i>R</i> ²	0.239	0.265	0.217	0.172

Note: With robust SEs and including covariates. Studies 1 and 4 are limited to the text treatment groups for comparability with Studies 2 and 5, which include only text treatments. In Study 2, we did not ask the trust in media questions for the second experiment assessing the oppositional rallying channel. Full table with covariates available as SM Table B16 and for the attentive subset of respondents as SM Table B22. Results are consistent for the attentive subset. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

whole? To answer this question, we examine whether claims of misinformation invoking informational uncertainty or oppositional rallying impact individuals' trust in media. Table 2 presents results that address the Trust in Media Hypothesis for Studies 1, 2, 4, and 5. We limit the samples to respondents in the text treatment groups due to the insignificant or weaker support impacts for video described above and to enable comparability across studies, as Studies 2 and 5 include only text treatments. However, SM Section B.11 also reports consistent null results for the impact of treatment on trust in media for video-based scandals.

Across studies, we find insignificant impacts on trust in media for both misinformation strategies. These results, while against our expectation, are consistent with recent findings showing that persuasive messages can move attitudes about the target of the message (in our case, the politician), but not other related attitudes (in our case, trust in media) (Coppock 2022, 36). It is possible that individuals' trust in media is already substantially low—indeed, only 28%, 33%, 23%, and 23% of individuals in control (exposed to a politician scandal, but not to any misinformation claims) in Studies 1, 2, 4, and 5, respectively, indicated that they trust the media.

While the results suggest that individuals' views about the media may not be affected by a single instance of misinformation about misinformation, future research should explore whether more prolonged or repeated exposure to politicians seeking liar's dividends can erode general trust in media. We return to trust in media when we compare claims of misinformation to alternative politician response strategies in the next section.

IS ALLEGING MISINFORMATION MORE EFFECTIVE THAN OTHER RESPONSES TO SCANDAL?

Results from Studies 1, 2, 4, and 5 indicate that politicians gain a liar's dividend when alleging misinformation, in response to text-based scandals, rather than remaining silent after scandal. Yet, remaining silent in

the face of scandal and attempting to allow a controversy to blow over is only one among several possible politician messaging strategies. It is possible that any active reply by a politician would proffer benefits compared to a non-response. To address this possibility, Study 3 compares claims of misinformation with two additional politician responses: a simple denial and an apology. These latter strategies have been studied previously as prominent types of politician reactions to transgressions (Gonzales et al. 1995) and have been found to mitigate reputational damage (Brenton 2011). Compared to these tried-and-true approaches, Study 3 allows us to assess whether claims of misinformation are effective in boosting support specifically *because* they invoke an environment saturated with misinformation.

For this study, we recruited 2,996 participants in October 2021 via Lucid. Participants were again randomly assigned to one of the four text-based politician scandals that we used in previous studies, followed, via random assignment, by one of three responses: the informational uncertainty rebuttal from prior studies, a simple denial that does not invoke misinformation, or an apology. We chose to use the informational uncertainty rebuttal (as opposed to oppositional rallying) because it most directly references the indirect harms to the informational environment due to misinformation.

We structured the denial and apology statements to be as similar as possible to the rebuttals claiming misinformation. The denial strategy, shown in Figure 7a, represents a flat-out denial common, for example, to politician sex scandals, such as Bill Clinton's infamous denial.

The apology statement provides an alternative in which the politician acknowledges the truth of the story and apologizes. This allows us to assess whether members of the public are more receptive to politicians who accept responsibility. The wording for this treatment, shown in Figure 7b, is based on the real reaction by John Murtha to the scandal used in our experiment.

FIGURE 7. Treatment Wordings for Alternative Politician Responses

[Politician Name] Denies that Events in Story Occurred

In response to the recent allegations, [Republican | Democrat] [Politician Name] firmly denied the story.

When asked about the incident, he said that it never occurred.

[Last Name] stated that “That did not happen. I never said that.”

(a) Simple Denial Treatment

[Politician Name] Acknowledges Story and Offers Apology

In response to the recent allegations, [Republican | Democrat] [Politician Name] acknowledged the story and apologized.

When asked about the incident, he said that it did occur.

[Last Name] stated that “Yes, I did say that, and I apologize for making those comments.”

(b) Apology Treatment

Rebuttals Invoking Misinformation Compared to Apologies and Simple Denials

Table 3 presents the results for Study 3 regarding the support gains that accrue to politicians from claiming misinformation or simply denying the scandal, along with impacts on belief in the scandal and trust in media. Unlike the prior studies, the comparison group is not politician non-response. Instead, claims of misinformation invoking informational uncertainty and simple denials are compared to an apology.³²

We find some evidence that claims of misinformation invoking informational uncertainty are more effective in generating politician support than alternative politician responses. In particular, claims of misinformation are more effective than apologizing (0.10 standard deviations, $p = 0.012$). In contrast, simple denials do not increase politician support relative to apologies ($p = 0.252$). While support gains from claims of misinformation are larger in magnitude than support gains due to more simple denials, the difference is not significant at conventional levels (0.06 standard deviations, $p = 0.167$). This finding could imply that a liar’s dividend in today’s informational environment is not meaningfully larger than dividends that would have accrued to liars in the past. Yet another possibility is that even simple denials are more effective in today’s informational ecosystem, a possibility that this study

³² While we originally pre-registered comparing allegations of misinformation to a pooled treatment group of apologies and simple denials together, we shift to comparing the effectiveness of claims on misinformation and simple denials against apologies based on reviewer and editor feedback. See SM Section A.5 for more details.

TABLE 3. Allegations of Misinformation and Simple Denials versus Apologies

	Support Index	Belief Index	Trust Index
	(1)	(2)	(3)
Info. Uncertain	0.103* (0.041)	-0.332*** (0.044)	-0.124** (0.040)
Simple Denial	0.047 (0.042)	-0.248*** (0.044)	-0.086* (0.040)
Sample	Study 3	Study 3	Study 3
N	2,994	2,994	2,994
R ²	0.105	0.067	0.223

Note: With robust SEs and including covariates. The reference group is composed of respondents who received an apology by the politician. Two respondents from the sample are excluded due to missing covariate information. Full table with covariates available as SM Table B17 and for the attentive subset of respondents as SM Table B23. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

cannot directly address. Additionally, while both apologies and simple denials are common politician responses to scandal, an open question is how rebuttals invoking misinformation compare to still *other* types of responses or variations of the treatments used here.

Table 3 also reveals that the liar’s dividend benefits to politicians are socially costly: they require deceiving the public, relative to the normatively preferable response of apologizing, creating the conditions for more uncertainty and less accountability. Belief in the scandal declines by a sizable 0.33 standard deviations when the public is misled in this way, likely because the apology response explicitly acknowledges the truth of the scandal. Trust in media also declines by 0.12 standard deviations. Simple denials also reduce belief in the scandal by 0.25 standard deviations and trust in media by 0.09 standard deviations. Compared to simple denials, claims of misinformation are marginally more effective in reducing public belief in a real scandal (-0.08 standard deviations, $p = 0.054$) but do not have a significantly different impact on trust in media ($p = 0.344$). Combined with the insignificant impacts on trust in media compared to politician non-response in Studies 1, 2, 4, and 5, it is possible that apologies *raise* trust in media relative to claims of misinformation and simple denials, as politicians corroborate media reporting on scandals through apologizing.³³

The results raise the concern that politicians can benefit by falsely alleging misinformation rather than taking responsibility for a scandal. This extends recent scholarship finding that politicians benefit more from denying scandals than from conceding and offering to take corrective action (Johnson 2018). Relatedly, while our results indicate that simple denials are not statistically more effective than apologies, *denials that employ the extra step of alleging misinformation* are indeed

³³ We thank an anonymous reviewer for raising this possibility.

more beneficial to politicians than apologies. In combination, these findings caution that political accountability in today's informational environment is especially difficult. Public figures may be incentivized to cry wolf over misinformation even when doing so undermines the foundation of accountability.

CONCLUSION

This study is the first to provide experimental evidence of the liar's dividend. We find that unscrupulous politicians willing to falsely claim misinformation may be rewarded with a reputational boost in the face of an otherwise damaging story. Claiming misinformation bolsters politician support more than remaining silent and allowing a scandal to blow over. It is also significantly more effective than apologizing—a preferable response for promoting trust and political accountability—and it is at least as effective as a more simple denial.

Interestingly, while the liar's dividend concept was originally developed in the context of concerns over the implications of deepfakes, we find that crying wolf about fake news may be more likely to pay off, while claims of deepfakes are, on the whole, unpersuasive. Scholars have debated the extent to which deepfakes are more believable and persuasive than text-based misinformation (Barari, Lucas, and Munger 2021; Wittenberg et al. 2021). Our results provide some evidence that *claims* of misinformation are differentially effective against text versus video scandals. In particular, attempts to discredit video using the informational uncertainty strategy appear to be ineffective. Yet results from November 2022 suggest that the oppositional rallying strategy may be effective even when politicians are confronted with video evidence of scandal. More research is warranted to evaluate whether this effect for video persists outside of election cycles or even increases as deepfakes become popularized, as they have in the public eye since late 2022.

Indeed, since administering our survey experiments, generative AI techniques such as diffusion-based models have become largely accessible to the public, with the ability to create diverse forms of inauthentic content through popular websites. By nearly eliminating the barrier to entry, it is now straightforward for regular citizens, politicians, and adversarial actors such as troll farms and hostile government actors to generate misinformation and disinformation en masse (Buchanan et al. 2021). These tools may also have benefits, such as simplifying political communication, enabling constituent outreach, and fostering political engagement through increased awareness and education. Yet they also have the potential to incentivize electoral interference and public manipulation, cheapen the value of human contact, and further undermine trust in political institutions. Despite attempts at increasing media literacy, detecting and labeling synthetic content, and regulating online platforms, it is not clear that these ameliorative strategies (Groh et al. 2022) are keeping pace with growing evidence of

harms. As a result, it is increasingly critical to better understand public and elite reactions to political misinformation.

Along these lines, this study also sought to better understand the strategies employed when politicians claim misinformation and the channels through which they affect individuals' attitudes. Drawing on real-world attempts to allege misinformation by public figures and relevant scholarly literature, this study proposes and evaluates two such strategies, which we term informational uncertainty and oppositional rallying. We find that both strategies are effective in raising politician support and that both are broadly appealing to co-partisans, independents, and even out-partisans. Therefore, further work is needed to explore why politicians employ these different strategies given that they seem to share certain features and impact similar audiences.

Moreover, we find evidence, although mixed and depending on the measure used, that both strategies may work, at least in part, through reducing respondents' belief that the events in the underlying scandal actually occurred. Given insignificant impacts on belief in Studies 1, 2, and 4, we nonetheless leave open the possibility of a belief-support disconnect in some individuals' processing of claims of misinformation and urge more extensive research on this topic. Finally, though Study 3 shows negative impacts on trust in media for claims of misinformation compared to apologies, null effects in other studies warrant further research into the relationship between claims of misinformation and trust in media, such as whether effects might be concentrated on particular news organizations or types of media.

Also, importantly, our study examines a certain type of political scandal, surrounding offensive comments largely related to race, ethnicity, gender, and identity. While some may consider these to be relatively minor gaffes, there are reasons to think that both making and responding to these kinds of comments is becoming more prevalent, such that this constitutes an important feature of modern political discourse worth understanding. An open question is whether other types of scandals, including potentially more severe ones, make the payout of the liar's dividend even greater. Additionally, for ethical reasons, this study is centered on inactive politicians and scandals that are not especially politically salient today. The liar's dividend may pay out even more for current political leaders, reinforced when political actors, organizations, and certain media sources act in concert to amplify misinformation and undermine trust. Yet, as members of the public may have stronger attachments and more information about currently active politicians, it is also possible that public attitudes may be more polarized and less malleable. Finally, while our study contrasted text and video, auditory misinformation without video has been shown to be slightly, if insignificantly, more credible than deepfake video (Barari, Lucas, and Munger 2021), and auditory information like vocal pitch is known to affect evaluations of politicians (Boussalis et al. 2021). Thus, a future

question concerns how auditory or visual information alone shapes the magnitude of the liar's dividend compared to audio-visual or textual information.

Overall, how concerned should we be about these findings? On the one hand, we find that false claims of misinformation are largely ineffective against video. This may suggest the benefits of using video evidence, whenever possible, when exposing the problematic behavior of politicians to promote political accountability. Moreover, in Study 2, we treated survey-takers with the informational uncertainty treatment and the informational uncertainty treatment followed by a fact-checking statement rejecting the politician's claim of misinformation. We found evidence, presented in SM Table A8, that fact-checking can eliminate liar's dividend support gains (reduction from a borderline-significant support gain of 0.08 standard deviations to an insignificant, negligible support difference relative to control). This is consistent with recent international research indicating that fact-checking is effective in combating misinformation (Porter and Wood 2021), which provides some hope for counteracting politician claims of misinformation.

Further, even if based on real scandals, our results established in an experimental context may overstate real-world effects. Our experimental design operates with near-simultaneous scandals and politician rebuttals, a "best-case scenario" which may allow for immediate counterbalancing of these competing signals (Chong and Druckman 2010), whereas in practice, repeated politician rebuttals or fact checks may occur later. While some research suggests individuals can recall their affective evaluations of politicians over time (Lodge, Steenbergen, and Brau 1995), effects of initial messages could also decay over time such that more recent politician rebuttals dominate. The role of timing, repetition, and countermessaging is important, and our design simplifies this more complex story.

On the other hand, many of our analyses are deliberately designed to be conservative. For example, we include inattentive respondents in our main results but find some larger effects for attentive participants. We also find larger effects when we omit more demanding support outcomes from our index. Even with a conservative approach, our study indicates potentially troubling consequences for political accountability. Might the indirect effects of misinformation be even more consequential for political accountability than the direct effects? Those political figures attempting to reap a liar's dividend may be counting on it.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <https://doi.org/10.1017/S0003055423001454>.

DATA AVAILABILITY STATEMENT

Research documentation, data, and replication code that support the findings of this study are openly

available at the American Political Science Review Dataverse: <https://doi.org/10.7910/DVN/MNO06W>.

ACKNOWLEDGMENTS

We are grateful to individuals at the Harvard Experimental Political Science Conference, participants in the annual meetings of the American Political Science Association and the Association for Public Policy Analysis and Management, and students in the Design & Analysis of Experiments course at Emory University for their comments and suggestions. We also thank Tom Clark, Adam Glynn, Zac Peskowitz, and Alex Coppock for feedback on this work.

FUNDING STATEMENT

We thank Emory University, Georgia Tech, Yale University, and Purdue University for providing funding for this project.

CONFLICT OF INTEREST

The authors declare no ethical issues or conflicts of interest in this research.

ETHICAL STANDARDS

The authors declare the human subjects research in this article was reviewed and approved by Emory University (Studies 1-5), the Georgia Institute of Technology (Studies 1-3), Yale University (Studies 4-5), and Purdue University (Studies 4-5) and certificate numbers are provided in the appendix. The authors affirm that this article adheres to the APSA's Principles and Guidance on Human Subject Research.

REFERENCES

- Arendt, Hannah. 1973. *The Origins of Totalitarianism*. New York: Harcourt, Brace, Jovanovich.
- Barari, Soubhik, Christopher Lucas, and Kevin Munger. 2021. "Political Deepfakes Are as Credible as Other Fake Media and (Sometimes) Real Media." Working Paper.
- Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58 (3): 739–53.
- Berinsky, Adam J., Michele F. Margolis, Michael W. Sances, and Christopher Warshaw. 2019. "Using Screeners to Measure Respondent Attention on Self-Administered Surveys: Which Items and How Many?" *Political Science Research and Methods* 9 (2): 430–7.
- Boussalis, Constantine, Travis G. Coan, Mirya R. Holman, and Stefan Müller. 2021. "Gender, Candidate Emotional Expression, and Voter Reactions During Televised Debates." *American Political Science Review* 115 (4): 1242–57.
- Bovet, Alexandre, and Hernán A. Makse. 2019. "Influence of Fake News in Twitter during the 2016 US Presidential Election." *Nature*

- Communications* 10 (7). <https://doi.org/10.1038/s41467-018-07761-2>
- Brenton, Scott. 2011. "When the Personal Becomes Political: Mitigating Damage Following Scandals." *Current Research in Social Psychology* 18, Article 4.
- Brutger, Ryan, Joshua D. Kertzer, Jonathan Renshon, Dustin Tingley, and Chagai M. Weiss. 2023. "Abstraction and Detail in Experimental Design." *American Journal of Political Science* 67 (4): 979–95.
- Buchanan, Ben, Andrew Lohn, Micah Musser, and Katerina Sedova. 2021. "Truth, Lies, and Automation: How Language Models Could Change Disinformation." Report. <https://cset.georgetown.edu/publication/truth-lies-and-automation/>.
- Bullock, John G., Alan S. Gerber, Seth J. Hill, and Gregory A. Huber. 2015. "Partisan Bias in Factual Beliefs about Politics." *Quarterly Journal of Political Science* 10 (4): 519–78.
- Chesney, Bobby, and Danielle Citron. 2019. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review* 107 (6): 1753–820.
- Chong, Dennis, and James N. Druckman. 2010. "Dynamic Public Opinion: Communication Effects over Time." *American Political Science Review* 104 (4): 663–80.
- Ciancaglini, Vincenzo, Craig Gibson, David Sancho, Odhran McCarthy, Maria Eira, Philipp Amann, and Aglika Klayn. 2020. "Malicious Uses and Abuses of Artificial Intelligence." Technical Report.
- Coppock, Alexander. 2022. *Persuasion in Parallel*. Chicago, IL: University of Chicago Press.
- Craig, Stephen C., and Paulina S. Cossette. 2022. "Eye of the Beholder: Partisanship, Identity, and the Politics of Sexual Harassment." *Political Behavior* 44: 749–77.
- Desai, Saoirse A. Connor, Toby D. Pilditch, and Jens K. Madsen. 2020. "The Rational Continued Influence of Misinformation." *Cognition* 205: 104453.
- Druckman, James N. 2012. "The Politics of Motivation." *Critical Review* 24 (2): 199–216.
- Erlanger, Steven. 2017. "'Fake News', Trump's Obsession, Is Now a Cudgel for Strongmen." *The New York Times*, December 12. <https://www.nytimes.com/2017/12/12/world/europe/trump-fake-news-dictators.html>.
- Flynn, D. J., Brendan Nyhan, and Jason Reifler. 2017. "The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs about Politics." *Political Psychology* 38 (S1): 127–50.
- Fowler, Anthony, and William G. Howell. 2023. "Updating amidst Disagreement: New Experimental Evidence on Partisan Cues." *Public Opinion Quarterly* 87 (1): 24–43.
- Funder, David C., and Daniel J. Ozer. 2019. "Evaluating Effect Size in Psychological Research: Sense and Nonsense." *Advances in Methods and Practices in Psychological Science* 2 (2): 156–68.
- Galston, William A. 2020. "Is Seeing Still Believing? The Deepfake Challenge to Truth in Politics." *Brookings*, January 8. <https://www.brookings.edu/research/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics/>.
- Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.
- Gonzales, Marti Hope, Margaret Bull Kovera, John L. Sullivan, and Virginia Chanley. 1995. "Private Reactions to Public Transgressions: Predictors of Evaluative Responses to Allegations of Political Misconduct." *Personality and Social Psychology Bulletin* 21 (2): 136–48.
- Graham, Matthew H. 2021. "Measuring Misperceptions?" *American Political Science Review* 117 (1): 80–102.
- Groh, Matthew, Ziv Epstein, Chaz Firestone, and Rosalind Picard. 2022. "Deepfake Detection by Human Crowds, Machines, and Machine-Informed Crowds." *Proceedings of the National Academy of Sciences* 119 (1): e2110013119.
- Guess, Andrew M., Brendan Nyhan, and Jason Reifler. 2020. "Exposure to Untrustworthy Websites in the 2016 US Election." *Nature Human Behaviour* 4 (5): 472–80.
- Hao, Karen. 2019. "The Biggest Threat of Deepfakes Isn't the Deepfakes Themselves." *MIT Technology Review*, October 10. <https://www.technologyreview.com/2019/10/10/132667/the-biggest-threat-of-deepfakes-isnt-the-deepfakes-themselves/>.
- Huang, Haifeng. 2015. "Propaganda as Signaling." *Comparative Politics* 47 (4): 419–44.
- Jerit, Jennifer, and Yangzi Zhao. 2020. "Political Misinformation." *Annual Review of Political Science* 23: 77–94.
- Johnson, Tyler. 2018. "Deny and Attack or Concede and Correct? Image Repair and the Politically Scandalized." *Journal of Political Marketing* 17 (3): 213–34.
- Kalla, Joshua L., and David E. Broockman. 2018. "The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments." *American Political Science Review* 112 (1): 148–66.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.
- Ladd, Jonathan M. 2011. *Why Americans Hate the Media and How It Matters*. Princeton, NJ: Princeton University Press.
- Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, et al. 2018. "The Science of Fake News." *Science* 359 (6380): 1094–6.
- Little, Andrew T. 2018. "Fake News, Propaganda, and Lies Can Be Pervasive Even If They Aren't Persuasive." *Critique* 11 (1): 21–34.
- Lodge, Milton, Marco R. Steenbergen, and Shawn Brau. 1995. "The Responsive Voter: Campaign Information and the Dynamics of Candidate Evaluation." *American Political Science Review* 89 (2): 309–26.
- Mitchell, Amy, Jeffrey Gottfried, Galen Stocking, Mason Walker, and Sophia Fedeli. 2019. "Many Americans Say Made-Up News Is a Critical Problem That Needs To Be Fixed." *Pew Research Center*, June 5. <https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/>.
- O'Shaughnessy, Nicholas Jackson. 2004. *Politics and Propaganda*. Manchester: Manchester University Press.
- Oppenheim, Maya. 2017. "Spanish Foreign Minister Claims Photos of Police Brutality Are 'Fake'." *The Independent*, October 2. <https://www.independent.co.uk/news/world/europe/catalan-independence-referendum-photos-police-violence-fake-a7978876.html>.
- Pennycook, Gordon, Ziv Epstein, Mohsen Mosleh, Antonio A. Arechar, Dean Eckles, and David G. Rand. 2021. "Shifting Attention to Accuracy Can Reduce Misinformation Online." *Nature* 592 (7855): 590–95.
- Peterson, Erik, and Shanto Iyengar. 2021. "Partisan Gaps in Political Information and Information-Seeking Behavior: Motivated Reasoning or Cheerleading?" *American Journal of Political Science* 65 (1): 133–47.
- Porter, Ethan, and Thomas J. Wood. 2021. "The Global Effectiveness of Fact-Checking: Evidence from Simultaneous Experiments in Argentina, Nigeria, South Africa, and the United Kingdom." *Proceedings of the National Academy of Sciences* 118 (37): e2104235118.
- Sabatier, Paul, Susan Hunter, and Susan McLaughlin. 1987. "The Devil Shift: Perceptions and Misperceptions of Opponents." *Western Political Quarterly* 40 (3): 449–76.
- Schiff, Kaylyn Jackson, Daniel S. Schiff, and Natália S. Bueno. 2024. "Replication Data for: The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability?" Harvard Dataverse. Dataset. <https://doi.org/10.7910/DVN/MNO06W>.
- Schwartz, Oscar. 2018. "You Thought Fake News Was Bad? Deep Fakes Are Where Truth Goes to Die." *The Guardian*, November 12. <https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth>.
- Singh, Shane P., and Judd R. Thornton. 2019. "Elections Activate Partisanship across Countries." *American Political Science Review* 113 (1): 248–53.
- Sundar, S. Shyam, Maria D. Molina, and Eugene Cho. 2021. "Seeing Is Believing: Is Video Modality More Powerful in Spreading Fake News via Online Messaging Apps?" *Journal of Computer-Mediated Communication* 26 (6): 301–19.
- Taber, Charles S., and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50 (3): 755–69.
- Tandoc, Edson C. Jr., Zheng Wei Lim, and Richard Ling. 2018. "Defining 'Fake News'." *Digital Journalism* 6 (2): 137–53.
- Tappin, Ben M., Adam J. Berinsky, and David G. Rand. 2023. "Partisans' Receptivity to Persuasive Messaging Is Undiminished by Countervailing Party Leader Cues." *Nature Human Behaviour* 7 (4): 568–82.

- Taylor, Shelley E., and Suzanne C. Thompson. 1982. "Stalking the Elusive 'Vividness' Effect." *Psychological Review* 89 (2): 155–81.
- Ternovski, John, Joshua Kalla, and Peter Aronow. 2022. "The Negative Consequences of Informing Voters about Deepfakes: Evidence from Two Survey Experiments." *Journal of Online Trust and Safety* 1 (2). <https://doi.org/10.54501/jots.v1i2.28>; <https://tsjournal.org/index.php/jots/article/view/28>.
- Toews, Rob. 2020. "Deepfakes Are Going to Wreak Havoc on Society. We Are Not Prepared." *Forbes*, May 25. <https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/>.
- Vaccari, Cristian, and Andrew Chadwick. 2020a. "'Deepfakes' Are Here. These Deceptive Videos Erode Trust in All News Media." *Washington Post*, May 28. <https://www.washingtonpost.com/politics/2020/05/28/deepfakes-are-here-these-deceptive-videos-erode-trust-all-news-media/>.
- Vaccari, Cristian, and Andrew Chadwick. 2020b. "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News." *Social Media + Society* 6 (1): 2056305120903408.
- Wang, Amy B. 2017. "A Mayor Denies It Is His Voice on Lewd, Racist Tapes. His Colleagues Say 'Resign.'" *Washington Post*, January 18. <https://www.washingtonpost.com/news/post-nation/wp/2017/01/17/a-mayor-denies-its-his-voice-on-lewd-racist-tapes-his-colleagues-say-resign/>.
- Watts, Duncan J., David M. Rothschild, and Markus Mobius. 2021. "Measuring the News and Its Impact on Democracy." *Proceedings of the National Academy of Sciences* 118 (15): e1912443118.
- West, Emily A., and Shanto Iyengar. 2020. "Partisanship as a Social Identity: Implications for Polarization." *Political Behavior* 44: 807–38.
- Wittenberg, Chloe, Ben M. Tappin, Adam J. Berinsky, and David G. Rand. 2021. "The (Minimal) Persuasive Advantage of Political Video over Text." *Proceedings of the National Academy of Sciences* 118 (47): e2114388118.