Research Article

Source alerts can reduce the harms of foreign disinformation

*Social media companies have begun to use content-based alerts in their efforts to combat mis- and disinformation, including fact-check corrections and warnings of possible falsity, such as "This claim about election fraud is disputed." Another harm reduction tool, source alerts, can be effective when a hidden foreign hand is known or suspected. This paper demonstrates that source alerts (e.g., **"Determined by Twitter to be a Russian government account"**) attached to pseudonymous posts can reduce the likelihood that users will believe and share political messages.*

Authors: Jason Ross Arnold (1), Alexandra Reckendorf (1), Amanda L. Wintersieck (1)
Affiliations: (1) Department of Political Science, Virginia Commonwealth University, USA

Research questions
- To what extent do source alerts affect social media users' tendency to believe and spread pseudonymous disinformation produced by foreign governments?
- Is there a difference between general (e.g., "foreign government") and specific (e.g., "Russian government") source alerts in mitigating the belief in and spread of disinformation?
- Are Democrats and Republicans equally likely to respond to source alerts that warn social media users about pseudonymous disinformation produced by foreign government sources?

Essay summary
- We conducted an experiment via Mechanical Turk (MTurk) (October 22-23, 2020; *N* = 1,483) exposing subjects to Facebook and Twitter memes based on real disinformation in order to test the efficacy of using source alerts to reduce the tendency of social media users to believe, like, share, and converse offline about pseudonymous disinformation related to the 2020 presidential election.
- We find that source alerts can reduce belief in the meme's claim and mitigate social media users' tendency to spread the disinformation online and offline, but the effects vary by the partisanship of the user, the type of social media, and the specificity of the alert.

- The findings advance our understanding of the theoretical promises and limitations of source alerts as a means of combating the belief in and spread of disinformation online. Specifically, our findings suggest that tech companies should continue to invest in efforts to track and warn users of disinformation and to expand their current practices to include country-of-origin tags. Our findings come short of suggesting a one-size-fits all solution, however, as we highlight the importance of considering partisanship as a confounding factor, as well as the importance of considering the nuances of competing social media environments.

Implications

Research on technological solutions to combat mis- and disinformation on social media platforms tends to focus on content-based warnings attached to messages that correct false claims or increase awareness of possible falsity (Bode & Vraga, 2015; Clayton et al., 2019; Lewandowsky et al., 2017; Margolin et al., 2018; Vraga & Bode, 2017). As a result, we know much less about the efficacy of source-based warnings, especially with regard to posts created by covert actors working for hostile foreign powers. Existing studies in this area have examined alerts attached to foreign governments' overt propaganda outlets, such as Russia's RT and Iran's PressTV platforms (Gleicher, 2020; Nassetta & Gross, 2020).

Foreign powers can have multiple persuasion goals related to covert psychological operations on social media, from changing users' attitudes about specific candidates or policies to decreasing trust in government or established facts. Some, like Russia, have long sought to "deepen the splits" – that is, to exacerbate existing socio-political conflicts (Osnos et al., 2017). Persuasion theorists since Aristotle have emphasized both source and message factors (among others) as key causes of attitude change (Aristotle, *Rhetoric*). Empirical research has identified multiple source characteristics that influence citizen decision-making, including: expert status (Clark et al., 2012; Hovland & Weiss, 1951; Hovland et al., 1953; Kelman & Hovland, 1953), occupation (Johnson et al., 2011), perceived ideology (Iyengar & Han, 2009; Stroud, 2011; Zaller, 1992), insider status (Carmines & Kuklinski, 1990), likeability (Chaiken, 1980; Lau & Redlawsk, 2001; Sniderman et al., 1991), party position or affiliation (Lodge & Hamill, 1986; Rahn, 1993), public approval (Mondak, 1993; Page et al., 1987), and trustworthiness (Aronson & Golden, 1962; Darmofal, 2005; Hovland & Weiss, 1951; Miller & Krosnick, 2000; Popkin, 1991). Source factors probably operate differently across media types (e.g., traditional versus social media), but we still require work that clarifies those differences.

Social media users may be less inclined to believe or share posts that are written pseudonymously by covert operatives of foreign powers if those individuals were made aware of the true source of the messages. Some users might regard a foreign power, such as the Russian government, as a deceptive, hostile force, while others might have mixed or indifferent views. Yet even those who do not see Russia as a malicious actor would likely think twice about believing or sharing a post if an alert made them aware of the underlying deception – for example, that a Russian agent pretended to be an American from Texas. A source-related alert could activate users' thoughts about the source's trustworthiness, likeability, and political affiliation, among others, which could halt the virality of a post that would otherwise be promoted through social media algorithms based on users' tendency to "like" or "share" the pseudonymous content. Overall, for multiple reasons, the identification of the foreign power behind a message will likely reduce its influence.

To determine whether, and to what extent, source alerts limit the influence of foreign disinformation, we conducted an experimental test of our arguments on a large national MTurk sample a week before the 2020 presidential election.

Our findings demonstrate the utility of an underused technological intervention that can reduce the belief in and spread of disinformation.[2] As more people continue to receive political information, actively or passively, through social media channels, the threat of pseudonymous disinformation infiltrating our discourse around elections, policy making, and current events remains high. In response, consumer and political pressures mount for social media companies to rein in disinformation and slow its spread. While we observed differences across partisanship, social media platforms, and alert structure variables, future research could clarify the conditions under which source alerts work best. For example, while Facebook users in our sample tended to show greater resistance to source alerts than Twitter users, other Facebook alerts might be more effective if they identify countries that users are predisposed to distrust. Just as our Facebook Democrats were moved to disbelieve and not share disinformation when it was tagged with a specific ("Russian government") source alert, which conforms to existing left-leaning concerns about Russian interference in general, Republican Facebook users might be more sensitive to alerts identifying China or Iran as the likely instigator. Future research should investigate whether Republican-leaning subjects are more responsive to source alerts that identify other countries. Future research might also clarify whether strength of partisanship is a confounding variable, especially as social media algorithms make it easier for disinformation to be targeted to strong partisans. Finally, future research should investigate the extent to which other political variables, such as strength of ideology, shape responses to alerts.
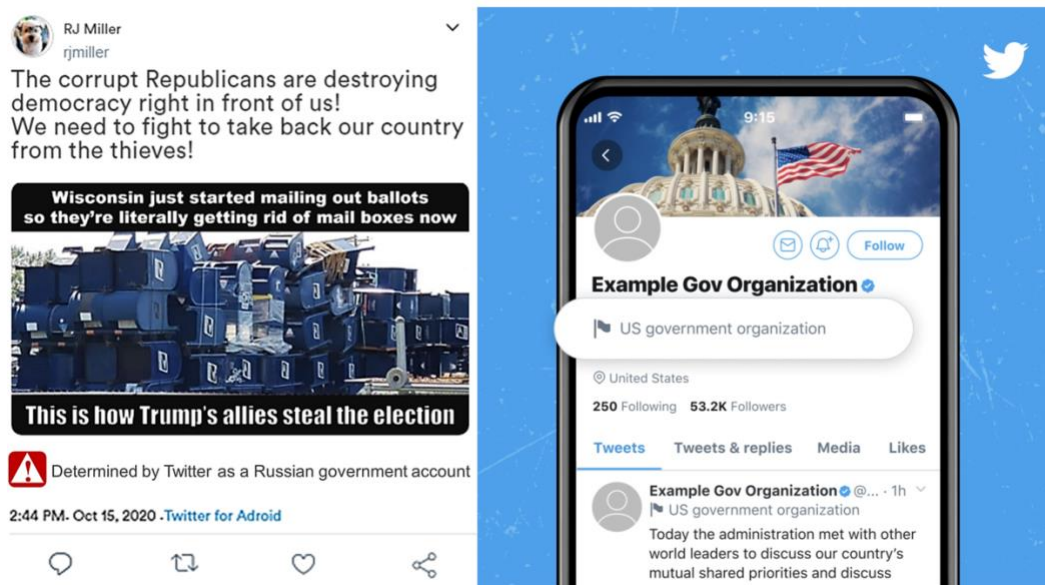

Figure 1. One of the 12 treatments used in the experiment and Twitter's example of an official foreign government account warning.

---

[2] We use the term *disinformation* throughout this article because our experimental manipulations are presented as originating from foreign governments' covert psychological operations. When a real-world disinformation post gains traction, it sometimes gets decoupled from the foreign power, at least directly – for instance, when an ordinary American citizen shares the message in a way that obscures the original poster. While that individual might not have the intentionality that is a key feature of disinformation campaigns, the message itself remains connected to the foreign campaign. Indeed, a goal of contemporary social media disinformation campaigns is to lead unsuspecting targets to spread the deceptive messages on their own. We follow Guess and Lyons (2020) in defining *misinformation* "as constituting a claim that contradicts or distorts common understandings of verifiable facts" (p. 10) and Tucker et al. (2018) in defining *disinformation* as a subset of misinformation that is "deliberately propagated false information" (p. 3, n. 1).

In light of our results, we recommend that Twitter, Facebook, and other social media companies experiment with a range of source alert formats for pseudonymous messages. The alerts shown to be effective in this study are similar to those introduced by Twitter in August 2020 "for government and state-affiliated media accounts." Firms might also experiment with more probabilistic language for accounts that are not clearly foreign government-controlled and that fail to cross a given threshold for "inauthentic behavior" (e.g., Facebook's Community Standards, Part IV, Section 20). Of course, because foreign powers try to conceal their efforts, tech companies will rarely be certain that an account is foreign-controlled. However, through human and AI-aided analyses, it is possible to estimate probabilities that accounts have links to foreign powers. In this study, we developed memes based on real disinformation that had been published on social media and had not been removed or otherwise targeted for violating community standards. Finally, we recommend that scholars continue to investigate the causes and consequences of the observed partisan, platform, and alert structure differences.

Findings

Finding 1: Source alerts do reduce social media users' tendency to believe pseudonymous disinformation in most social media environments; however, no significant effect is found for Republican or Democratic Facebook users exposed to the general ("foreign government") source alert.

The effects of Table 1 are depicted as predicted probabilities in Figure 1. Looking first at the Facebook conditions, exposure to the Russia alert reduced the probability that a Democratic respondent said the claims were truthful by 18.6%, compared to the no alert condition. Among Democrats in the Twitter conditions, exposure to the general foreign alert reduced the probability that a respondent said the claims were truthful by 26.4%, and exposure to the Russia alert reduced the probability that a respondent said the claims were truthful by 23.2%, compared to the no alert condition. Finally, among Republicans in the Twitter conditions, exposure to the foreign alert reduced the probability that a respondent said the claims were truthful by 16.8%, and exposure to the Russia alert reduced the probability that a respondent said the claims were truthful by 8.3%, compared to the no alert condition. Results were not statistically significant for Republicans in the Facebook conditions.
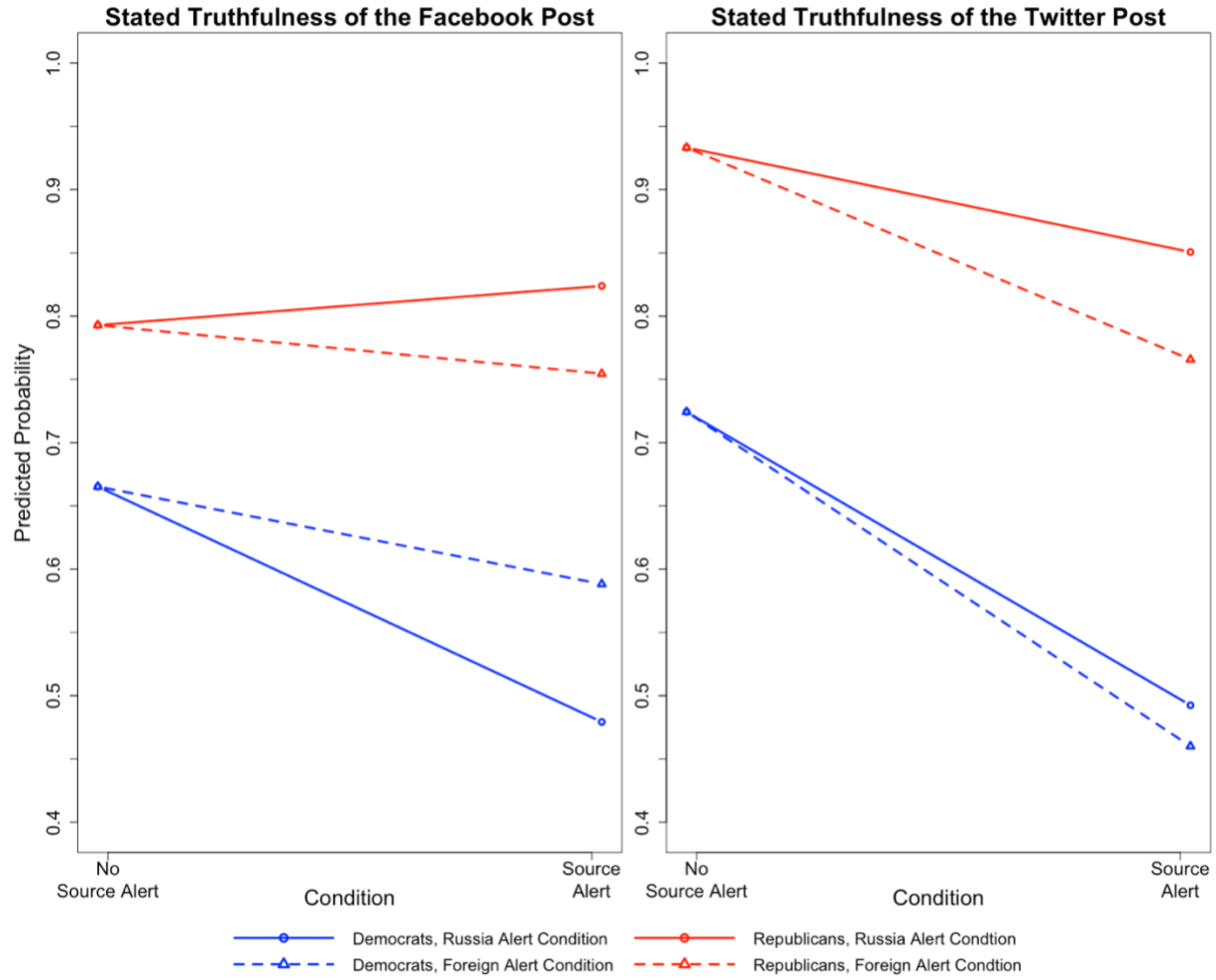
Figure 2. *Probability that respondent stated that the information in the social media post was truthful.[3]*

---

[3] For Figure 1, age, education, share political posts, social media use, and level of authoritarianism are held at the mean. Gender equals female. South equals zero.

Table 1. *GLM, agreement with the truthfulness of the information in the social media post.*

	Facebook Conditions		Twitter Conditions	
	Democrats	Republicans	Democrats	Republicans
Intercept	-1.06 (0.78)	-0.88 (0.82)	-0.72 (1.08)	0.37 (1.44)
Facebook Foreign	-0.33 (0.27)	0.22 (0.27)	--	--
Facebook Russia	-0.77*** (0.26)	0.2 (0.29)	--	--
Twitter Foreign	--	--	-1.13*** (0.35)	-1.45*** (0.53)
Twitter Russia	--	--	-0.99*** (0.36)	-0.89* (0.47)
Age	-0.02** (0.01)	-0.02* (0.01)	-0.01 (0.01)	0.02 (0.02)
Education	-0.03 (0.09)	-0.05 (0.1)	-0.04 (0.13)	-0.09 (0.15)
Female	-0.44** (0.22)	0.23 (0.23)	0.19 (0.29)	1.16** (0.47)
South	0.13 (0.22)	0.24 (0.23)	0.11 (0.29)	-0.14 (0.39)
Share Political Information	0.92*** (0.13)	0.93*** (0.13)	0.49*** (0.16)	0.68*** (0.22)
Social Media Use	0.01 (0.1)	0.03 (0.12)	0.06 (0.16)	-0.29 (0.19)
Authoritarianism	0.39*** (0.08)	0.09 (0.09)	0.27** (0.11)	0.08 (0.14)
AIC	549.18	503.54	329.71	192.51
N	493	503	248	149

*Note: The dependent variable is subjects' agreement with the truthfulness of the social media post, coded so that 0=False, 1=True. Twitter Foreign, Twitter Russia, Facebook Foreign, Facebook Russia are dummy variables where 0 indicates that the respondent was not in the condition and 1 indicates the respondent was in the source alert condition. Standard errors are in parentheses. The model is a logit. *** $p < .01$ **$p < .05$ *$p < .10$*

Finding 2: Source alerts do reduce the spread of disinformation on Twitter regardless of the users' partisanship; however, source alerts do not reduce the spread of disinformation online on Facebook.

In all but one experimental condition, Democratic and Republican Twitter users alike reported that they would be less likely to "like" and "retweet" the disinformation contained in our meme after being exposed to either the general ("foreign government") or specific ("Russian government") source alert treatment. As Table 2 shows, for Democrats on Twitter, both the general (p = .039) and specific (p = .008) treatments produced a reported reduction in the likelihood that users would "like" the disinformation. Those in the

specific treatment also reported a reduction in the likelihood of sharing the disinformation by retweeting it (p = .016). For Republicans on Twitter, the general alert was effective in reducing the tendency of liking (p = .000) and retweeting (p = .011) the disinformation, and the specific alert was effective in reducing the tendency of liking (p = .004) and retweeting (p = .013) the disinformation.

The findings for Twitter users are notably different from the Facebook users. As Table 2 shows, the tendency of both Democratic and Republican Facebook users to "like" and "share" our experimental memes remained unmoved by our general and specific source alert treatments. The difference in the efficacy of source alerts between the two platforms might be due to the different ways users choose to engage with and exchange information within those social media environments, based on the nuances of those platforms and their networks. This is an area for future scholarly research.

Table 2. *Likelihood that subjects would like or share/retweet social media disinformation on Facebook and Twitter.*

	LIKE				SHARE/RETWEET			
	Facebook Conditions		Twitter Conditions		Facebook Conditions		Twitter Conditions	
	Democrats	Republicans	Democrats	Republicans	Democrats	Republicans	Democrats	Republicans
Facebook Foreign	-0.26 (0.21)	-0.1 (0.20)	--	--	-0.13 (0.21)	0.11 (0.2)	--	--
Facebook Russia	-0.24 (0.2)	0.01 (0.20)	--	--	-0.21 (0.2)	0.08 (0.21)	--	--
Twitter Foreign	--	--	-0.58** (0.29)	-1.44*** (0.4)	--	--	-0.31 (0.3)	-1.03** (0.4)
Twitter Russia	--	--	-0.79*** (0.3)	-1.03*** (0.36)	--	--	-0.74** (0.31)	-0.88** (0.35)
Age	-0.03** (0.01)	-0.02*** (0.01)	0 (0.01)	0.01 (0.01)	-0.03*** (0.01)	-0.02*** (0.01)	0 (0.01)	0 (0.01)
Education	-0.07 (0.07)	-0.03 (0.07)	0.03 (0.11)	-0.03 (0.12)	-0.14* (0.08)	-0.1 (0.07)	0.07 (0.12)	-0.11 (0.12)
Female	-0.57*** (0.17)	0.13 (0.17)	0.31 (0.25)	0.77** (0.33)	-0.68*** (0.18)	-0.02 (0.17)	0.36 (0.25)	0.39 (0.33)
South	0.17 (0.17)	-0.05 (0.17)	-0.54** (0.25)	-0.12 (0.3)	0.11 (0.18)	0.02 (0.17)	-0.55** (0.26)	-0.03 (0.31)
Share Political Information	1.39*** (0.12)	1.43*** (0.11)	0.78*** (0.15)	1.39*** (0.2)	1.47*** (0.13)	1.73*** (0.12)	1.1*** (0.16)	1.63*** (0.22)
Social Media Use	0.11 (0.08)	0.07 (0.09)	0.17 (0.14)	-0.30** (0.13)	0.12 (0.09)	0.13 (0.09)	0.01 (0.15)	-0.17 (0.13)
Authoritarianism	0.19*** (0.06)	-0.02 (0.07)	0.51*** (0.1)	0.28** (0.11)	0.34*** (0.06)	0.07 (0.07)	0.52*** (0.1)	0.31*** (0.12)
AIC	1573.43	1555.67	828.37	520.039	1483.59	1532.58	761.68	515.34
N	493	503	248	160	493	503	248	160

*Note: The dependent variables are subjects' indication of their likelihood to like or share/retweet the social media post. Options ranged from -3 (Very Unlikely) to 3 (Very Likely). Twitter Foreign, Twitter Russia, Facebook Foreign, Facebook Russia are dummy variables where 0 indicates that the respondent was not in the condition and 1 indicates the respondent was in the source alert condition. Standard errors are in parentheses. Cut points excluded for brevity. The model is an ordered logit. *** p < .01 **p < .05 *p < .10*

Finding 3: Source alerts reduce the spread of disinformation offline among Democrats in the Facebook conditions; however, the effect for Republicans is more limited and the observed effect is less generalizable to Twitter.

Source alerts reduced the spread of offline disinformation by Democratic Facebook users. We observed no such effects in the Republican Facebook conditions. Democratic Facebook users in the specific ("Russian government") treatment reported a reduced likelihood of both initiating conversation about the disinformation presented in the meme (p = .046) and of engaging in conversation about the disinformation with friends, family members, acquaintances, etc. (p = .042) (see Table 3). Democratic Facebook users also reported a significantly reduced likelihood of initiating conversation when exposed to the general ("foreign government") treatment (p = .013). Thus, Democratic Facebook users were relatively consistent in responding to source alerts, which reduced the likelihood that they would spread disinformation offline. In contrast, only the general source alert had a statistically significant effect in the Democratic Twitter condition (p = .040) on the tendency to initiate conversations. Only the general source alert had a statistically significant effect in the Republican Twitter condition (p = .020) on the tendency to engage in conversation about the information in the meme. Our inconsistent Twitter results limit our ability to draw general conclusions about that platform. However, we did find statistically significant results for Democratic and Republican Twitter users related to the dependent variable about initiating offline conversations. That alone provides justification for continuing to explore the efficacy of source cues on Twitter.

Table 3. *Likelihood that subjects would initiate or otherwise engage in offline conversations about the social media posts.*

	INITIATE OFFLINE CONVERSATION				ENGAGE IN OFFLINE CONVERSATION			
	Facebook Conditions		Twitter Conditions		Facebook Conditions		Twitter Conditions	
	Democrats	Republicans	Democrats	Republicans	Democrats	Republicans	Democrats	Republicans
Facebook Foreign	-0.51** (0.2)	-0.2 (0.2)	--	--	-0.14 (0.2)	-0.13 (0.2)	--	--
Facebook Russia	-0.4** (0.2)	-0.09 (0.2)	--	--	-0.4** (0.2)	-0.01 (0.2)	--	--
Twitter Foreign	--	--	-0.59** (0.29)	-0.73* (0.38)	--	--	-0.45 (0.28)	-0.88** (0.38)
Twitter Russia	--	--	-0.53* (0.29)	-0.1 (0.34)	--	--	-0.3 (0.29)	-0.18 (0.35)
Age	-0.02** (0.01)	-0.01* (0.01)	0 (0.01)	0.01 (0.01)	-0.01* (0.01)	-0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Education	-0.04 (0.07)	0 (0.07)	0.16 (0.11)	-0.08 (0.11)	0.05 (0.07)	-0.09 (0.1)	0.16 (0.11)	0.01 (0.11)
Female	-0.43*** (0.17)	0.04 (0.16)	0.63*** (0.24)	0.46 (0.31)	-0.3* (0.17)	0.19 (0.16)	0.69*** (0.24)	0.29 (0.32)
South	0.02 (0.17)	0.01 (0.16)	-0.51** (0.24)	-0.22 (0.29)	0.11 (0.17)	-0.12 (0.16)	-0.34 (0.24)	-0.26 (0.29)
Share Political Information	1.49*** (0.12)	1.23*** (0.11)	0.69*** (0.14)	1.23*** (0.2)	1.24*** (0.11)	1.19*** (0.11)	0.58*** (0.14)	1.26*** (0.19)
Social Media Use	0.11 (0.08)	0.14* (0.09)	0.17 (0.14)	-0.08 (0.13)	0.09 (0.08)	0.16* (0.08)	0.1 (0.14)	-0.05 (0.13)
Authoritarianism	0.12** (0.06)	0.00 (0.06)	0.33*** (0.09)	0.09 (0.11)	0.13** (0.06)	0.02 (0.06)	0.3*** (0.09)	0.18 (0.11)
AIC	1641.51	1669.26	890.85	576.81	1662.26	1654.31	903.71	568.782
N	493	503	248	160	493	503	248	160

Note: The dependent variables are subjects' indication of their likelihood to initiate or engage in an offline conversation about the information in the post. Options ranged from -3 (Very Unlikely) to 3 (Very Likely). Twitter Foreign, Twitter Russia, Facebook Foreign, Facebook Russia are dummy variables where 0 indicates that the respondent was not in the condition and 1 indicates the respondent was in the source alert condition. Standard errors are in parentheses. Cut points excluded for brevity. The model is an ordered logit. *** p < .01 **p < .05 *p < .10

Finding 4: Democratic and Republican partisans respond differently to source alerts, but the difference is conditioned by users' preferred social media platform.

Although some differences do emerge between Democratic and Republican responses to source alerts, the effects of our treatments on Democratic and Republican Twitter users generally mirrored one another. Furthermore, Democrats on Facebook acted more like Democrats and Republicans on Twitter, when we analyzed the effectiveness of source alerts on the tendency to believe disinformation and to spread disinformation online. However, they responded similarly to Republicans on Facebook, when we analyzed source alerts' ineffectiveness to curb their tendency to spread disinformation online using Facebook's "like" and "share" features, perhaps due to subtle differences in how Facebook users of both parties engage with information on the platform. Most notably, the Republican Facebook users appeared resistant to source alerts, a pattern that distinguishes them not only from Democratic Facebook users, but Twitter users of both parties. There appears to be something distinct about Republicans who prefer Facebook as a platform for their social communications or how they engage with that specific social media environment. This distinction could cause them to continue to perpetuate disinformation online and offline regardless of the company's attempts to put source alerts in place to warn of pseudonymous foreign disinformation attempts. Future research should focus on this problem.

Methods

To determine whether, and to what extent, source alerts limit the influence of foreign disinformation, we conducted an experimental test of the following hypotheses on a large national MTurk sample a week before the 2020 presidential election:

Hypothesis$_1$: Exposure to source alerts will reduce social media users' tendency to believe pseudonymous disinformation.

Hypothesis$_2$: Exposure to source alerts will mitigate social media users' tendency to spread pseudonymous disinformation online.
H$_{2a}$: Exposure to source alerts will mitigate social media users' tendency to "like" pseudonymous disinformation.
H$_{2b}$: Exposure to source alerts will mitigate social media users' tendency to "share/retweet" pseudonymous disinformation.

Hypothesis$_3$: Exposure to source alerts will mitigate social media users' tendency to spread pseudonymous disinformation offline.
H$_{3a}$: Exposure to source alerts will mitigate social media users' tendency to initiate conversations about pseudonymous disinformation.
H$_{3b}$: Exposure to source alerts will mitigate social media users' tendency to engage in conversations about pseudonymous disinformation.

Data collection

In order to test our hypotheses, we conducted an experiment utilizing disinformation directly related to the 2020 U.S. presidential election. The experiment took place on October 22-23, 2020 (*N* = 1,483) and is a two (social media platforms) by two (party consistent message) by two (source cue) design, producing eight experimental conditions and four control conditions. Subjects were recruited via MTurk. All subjects

were required to be at least 18 years of age and U.S. citizens. The survey was built and distributed via SurveyMonkey and randomization was used.[4]

Prior to exposure to the treatments, we asked subjects about their social media usage and party identification. These questions were used to branch subjects into the appropriate stimuli groups. That is, subjects who indicated a preference for Twitter were branched into the Twitter treatment group; while subjects who indicated a preference for Facebook received a treatment that appeared to be from Facebook. The Twitter and the Facebook posts use the same images, wording, and user profile details, but the meme was altered to appear as a Facebook or Twitter post. Additionally, Democrats and Republicans were exposed to treatment memes that were relevant to their stated partisanship. Democrats were shown a message about alleged voter suppression efforts by Republicans, while Republicans were shown a message about voter fraud efforts allegedly perpetrated by Democrats.[5] After the filter questions, all subjects were randomly assigned to the source cues (control, Russian Government Account, Foreign Government Account). Source alerts were presented adjacent to the posts with a highly visible cautionary symbol to draw subjects' attention to the alert. Finally, subjects were asked a number of questions about the social media post they viewed and how they would engage with the post if it appeared on their Facebook or Twitter feed.

Overall, our MTurk sample is more male, educated, and younger than the U.S. population: 42.2% are female, 69.7% have post-secondary degrees, and 24.8% are 18-29 years old, 59.87% are 30-49, and 15.3% are 50 or older.[6] In addition, 77.6% are white, while 39% reside in the south. Moreover, 52.1% side with the Democratic Party, with 47.9% noting a closer link with the Republican Party.

---

[4] We conducted randomization tests involving partisanship and social media types. Results of the randomization checks are available from the authors. Thirteen subjects dropped out of the experiment. The experiment was estimated to take approximately seven minutes to complete.

[5] In response to the partisanship question, subjects who initially indicated that they were *independent* or *something else* were then prompted to indicate whether they think of themselves as being closer to the Republican or Democratic Party and branched into an experimental condition accordingly. Self-proclaimed Independents and non-partisans often act in the same fashion as their partisan counterparts, marking partisan independence as a matter of self-presentation rather than actual beliefs and behaviors (Petrocik, 2009). Additionally, negative partisanship has been identified as a primary motivator in the way Americans respond to political parties and candidates (Abramowitz & Webster, 2016); our treatments capture appeals to negative partisanship more than they capture loyalty to any one particular party.

[6] There is an emerging consensus among political methodologists about the efficacy of MTurk sampling procedures in experimental research. Berinsky et al. (2012), for example, find MTurk samples to be more representative than "the modal sample in published experimental political science" (p. 351) even if it is less representative than national probability samples or internet-based panels.
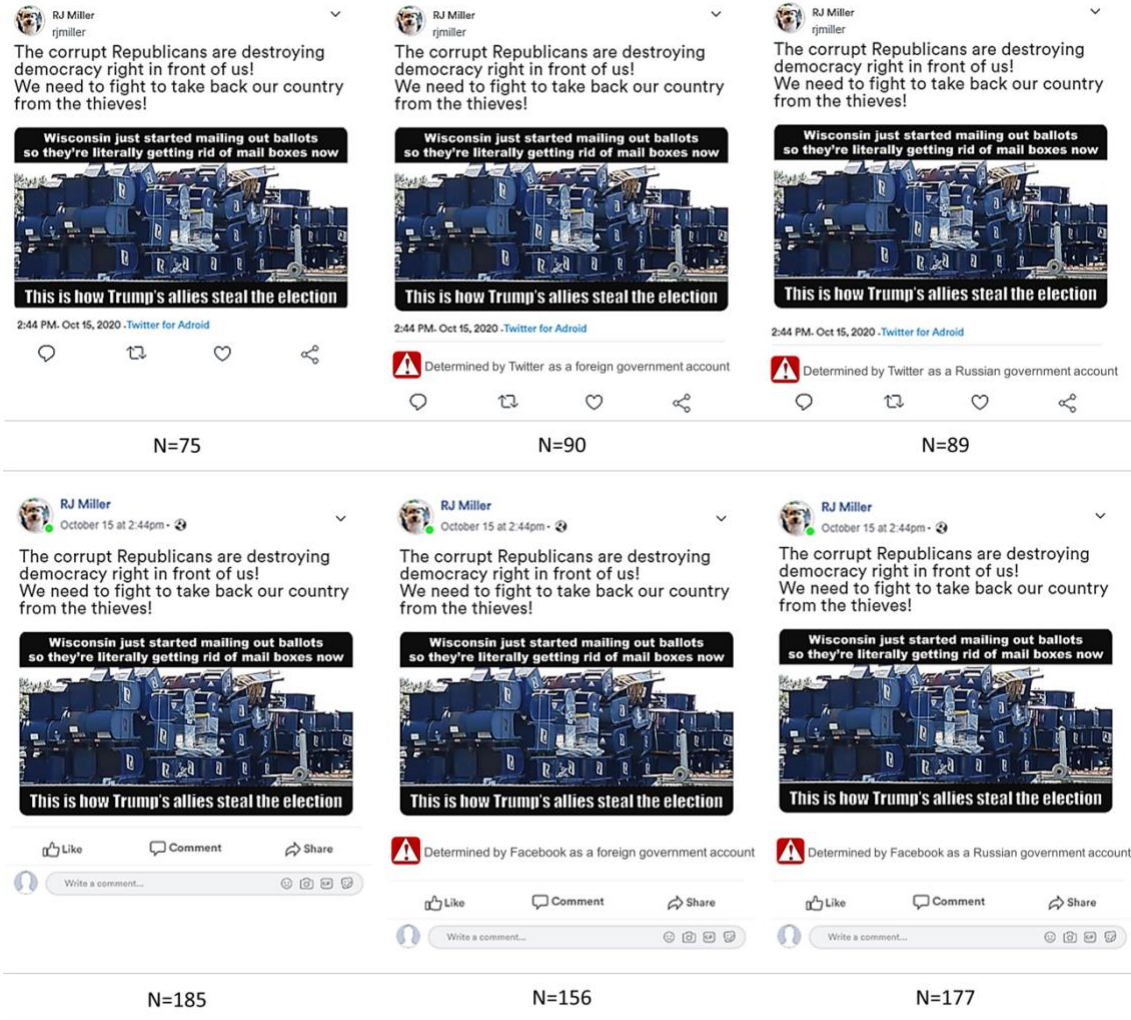
Figure 3. Disinformation post shown to Democrats, modified by social media type and treatment group.

Figure 4. *Disinformation post shown to Republicans, modified by social media type and treatment group.*[7]

Dependent variables

Subjects answered a number of post-treatment response items to examine their perceptions and behavioral intentions. Specifically, we asked respondents to report how truthful the information in the post was (0 = false and 1 = true), how likely they would be to "like" and "share"/"retweet" the post, and how likely they would be to initiate conversation and engage in conversation about the post offline (each coded -3 to 3, ranging from least to most likely). All of those items, along with others measuring respondents' partisanship and social media habits, are included in the appendix.

---

[7] While we acknowledge that fewer subjects chose Twitter than Facebook, there are enough subjects in each Twitter condition for statistical power.

Independent variables

In addition to the condition dummy variables, we include a number of control variables in our models. Age is included because the mean age of subjects who chose the Twitter conditions was lower than the mean age of subjects who chose Facebook. We include a measure of gender because the exclusion of 13 incomplete responses resulted in a gender imbalance in two of our Twitter conditions. Education is included because belief in conspiracy thinking is more common among less educated individuals (Goertzel, 1994). We include a variable for the South because regional differences are common in political behavioral research. Subjects' tendency to share political information and their rate of social media use are used to control for social media habits. Finally, we included an authoritarian personality measure because authoritarianism has been shown to be a key variable in influencing political behaviors in multiple domains (Feldman, 2003). Specifically, more authoritarian individuals have been found to staunchly defend prior attitudes and beliefs. Low authoritarian individuals, on the other hand, have demonstrated a greater need for cognition (Hetherington & Weiler, 2009; Lavine et al., 2005; Wintersieck, forthcoming). As a result, high authoritarians may be less moved by source alerts on information they are likely to believe, while low authoritarians may be more likely to take up these cues when making assessments about the memes. Information about coding these variables is presented in the appendix.

External validity

A number of steps were taken to increase the external validity of the experiment. First, we designed memes based on real disinformation about election fraud. Each meme was similar in imagery, language, and length. Furthermore, each meme presented disinformation that sows discord in U.S. electoral politics, a goal that would further the interests of foreign governments, like Russia, and that could, therefore, conceivably be spread by those pseudonymous sources. Second, we used two filter questions prior to exposure to the treatment to mimic actual social media engagement for the subjects. Specifically, we used a party ID filter that allowed us to create a typical social media echo chamber by filtering Democrats into conditions with misinformation that a Democrat would likely encounter on social media and Republicans into conditions with misinformation that a Republican would likely encounter on social media. Additionally, the social media filter allowed us to put subjects into social media treatment conditions with social media platforms that they were more likely to be familiar with. That is, if a subject does not use Twitter, or understand the Twitter platform, it is highly unlikely that they would indicate a willingness to share the tweet, etc. This becomes a confounding variable because we are unable to parse out whether lack of willingness to share is because of a lack of understanding of the platform or because of the source cue.

Bibliography

Abramowitz, A. I., & Webster, S. (2016). The rise of negative partisanship and the nationalization of U.S. elections in the 21st century. *Electoral Studies*, *41*, 12–22. https://doi.org/10.1016/j.electstud.2015.11.001

Aronson, E., & Golden, B. W. (1962). The effect of relevant and irrelevant aspects of communicator credibility on opinion change. *Journal of Personality, 30*(2). https://doi.org/10.1111/j.1467-6494.1962.tb01680.x

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368. https://doi.org/10.1093/pan/mpr057

Bode, L., & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication*, *65*(4), 619–638. https://doi.org/10.1111/jcom.12166

Carmines, E. G., & Kuklinski, J. H. (1990). Incentives, opportunities, and the logic of public opinion in American political representation. In J. A. Ferejohn & J. H. Kuklinski (Eds.), *Information and democratic processes* (pp. 240-68). University of Illinois Press.

Chaiken, R. L. (1977). *The use of source versus message cues in persuasion: an information processing analysis* (Publication No. 1656) [Doctoral dissertation, University of Massachusetts Amherst]. https://scholarworks.umass.edu/dissertations_1/1656

Clark, J. K., Wegener, D. T., Habashi, M. M., & Evans, A. T. (2012). Source expertise and persuasion: The effects of perceived opposition or support on message scrutiny. *Personality and Social Psychology Bulletin*, *38*(1), 90–100. https://doi.org/10.1177/0146167211420733

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R. Scholz-Bright, R., Welch, A., Wolff, A. G., Zhou, A. & Nyhan, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, *42*(4), 1073–1095. https://doi.org/10.1007/s11109-019-09533-0

Darmofal, D. (2005). Elite cues and citizen disagreement with expert opinion. *Political Research Quarterly*, *58*(3), 381–395. https://doi.org/10.2307/3595609

Feldman, S. (2003). Enforcing social conformity: A theory of authoritarianism. *Political Psychology*, *24*(1), 41–74. https://doi.org/10.1111/0162-895x.00316

Gleicher, N. (2020, June 4). *Labeling state-controlled media on Facebook*. Facebook Newsroom. https://about.fb.com/news/2020/06/labeling-state-controlled-media/

Guess, A. M., & Lyons, B. A. (2020) Misinformation, disinformation, and online propaganda. In N. Persily & J. Tucker (Eds.), *Social media and democracy: the state of the field, prospects for reform* (pp. 10–33). Cambridge University Press.

Hetherington, M. J., & Weiler, J. D. (2009). *Authoritarianism and polarization in American politics*. Cambridge University Press. https://doi.org/10.1017/cbo9780511802331

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, *15*(4), 635–650. https://doi.org/10.1086/266350

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion*. Yale University Press.

Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, *59*(1), 19–39. https://doi.org/10.1111/j.1460-2466.2008.01402.x

Johnson, T., Dunaway, J., & Weber, C. R. (2011). Consider the source: Variations in the effects of negative campaign messages. *Journal of Integrated Social Sciences*, *2*(1), 98–127. https://www.jiss.org/documents/volume_2/issue_1/JISS_2011_Consider_the_Source.pdf

Kelman, H. C., & Hovland, C. I. (1953). "Reinstatement" of the communicator in delayed measurement of opinion change. *The Journal of Abnormal and Social Psychology*, *48*(3), 327. https://doi.org/10.1037/h0061861

Lau, R. R., & Redlawsk, D. P. (2001). Advantages and disadvantages of cognitive heuristics in political decision making. *American Journal of Political Science 45*(4), 951–971. https://doi.org/10.2307/2669334

Lavine, H., Lodge, M., & Freitas, K. (2005). Threat, authoritarianism, and selective exposure to information. *Political Psychology*, *26*(2), 219–244. https://doi.org/10.1111/j.1467-9221.2005.00416.x

Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

Lodge, M., & Hamill, R. (1986). A partisan schema for political information processing. *The American Political Science Review*, *80*(2), 505–520. https://doi.org/10.2307/1958271

Margolin, D. B., Hannak, A., & Weber, I. (2018). Political fact-checking on Twitter: When do corrections have an effect? *Political Communication*, *35*(2), 196–219. https://doi.org/10.1080/10584609.2017.1334018

Miller, J. M., & Krosnick, J. A. (2000). News media impact on the ingredients of presidential evaluations: Politically knowledgeable citizens are guided by a trusted source. *American Journal of Political Science, 44*(2), 301–315. https://doi.org/10.2307/2669312

Mondak, J. J. (1993). Source cues and policy approval: The cognitive dynamics of public support for the Reagan agenda. *American Journal of Political Science*, *37*(1), 186–212. https://doi.org/10.2307/2111529

Nassetta, J., & Gross, K. (2020). State media warning labels can counteract the effects of foreign misinformation. *Harvard Kennedy School (HKS) Misinformation Review, 1*(7). https://doi.org/10.37016/mr-2020-45

Osnos, E., Remnick, D., & Yaffa, J. (2017, March 6). Trump, Putin, and the new cold war. *The New Yorker*. https://www.newyorker.com/magazine/2017/03/06/trump-putin-and-the-new-cold-war

Page, B. I., Shapiro, R. Y., & Dempsey, G. R. (1987). What moves public opinion? *The American Political Science Review*, *81*(1), 23–43. https://doi.org/10.2307/1960777

Petrocik, J. R. (2009). Measuring party support: Leaners are not independents. *Electoral Studies*, *28*(4), 562–572. https://doi.org/10.1016/j.electstud.2009.05.022

Popkin, S. L. (1991). *The reasoning voter: Communication and persuasion in presidential campaigns*. University of Chicago Press.

Rahn, W. M. (1993). The role of partisan stereotypes in information processing about political candidates. *American Journal of Political Science*, *37*(2), 472–496. https://doi.org/10.2307/2111381

Sniderman, P. M., Brody, R. A., & Tetlock, P. E. (1991). *Reasoning and choice: Explorations in political psychology*. Cambridge University Press.

Stroud, N. J. (2011). *Niche news: The politics of news choice*. Oxford University Press.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018, March 29). Social media, political polarization, and political disinformation: A review of the scientific literature. Hewlett Foundation. https://www.hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf

Vraga, E. K., & Bode, L. (2017). Using expert sources to correct health misinformation in social media. *Science Communication*, *39*(5), 621–645. https://doi.org/10.1177/1075547017731776

Wintersieck, A. (forthcoming). Authoritarianism, fact-checking, and citizens' response to presidential election information. In Barker, D., & Suhay, E. Elizabeth (Eds.), *The politics of truth in polarized America.* Oxford University Press.

Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge University Press.

Ethics
This study was approved on October 22, 2020, by the Institutional Review Board at Virginia Commonwealth University (HM20017649_Ame1). All participants provided informed consent prior to participating in these studies.

The authors determined gender categories in the survey with a set of three questions. The categories included: "other," "male," "female," "transgender," "do not identify as male, female, or transgender," and "prefer not to say." Respondents who chose "other" were then given an open-ended question. However, all subjects but one selected "male" or "female." Moreover, because of the need to drop 13 cases who did not complete the experiment, there was a gender imbalance in the Democrat Facebook condition. As a result, the authors decided to control for gender in the statistical models, using a binary measure. The authors included gender in the study because it is a key cause of political behavior, as are race and ethnic identities. Although the authors did not use race and ethnicity in the statistical models, they did ask respondents about their racial and ethnic identities, using questions from NORC's AmeriSpeak battery.

Appendix A: Experimental conditions and subjects

Appendix Table A. *Experimental conditions and subjects.*

Condition	N
Facebook Republican Control	182
Facebook Republican Russia Alert	173
Facebook Republican Foreign Alert	184
Facebook Democratic Control	185
Facebook Democratic Russia Alert	177
Facebook Democratic Foreign Alert	156
Twitter Republican Control	54
Twitter Republican Russia Alert	66
Twitter Republican Foreign Alert	52
Twitter Democratic Control	75
Twitter Democratic Russia Alert	89
Twitter Democratic Foreign Alert	90

Appendix B: Survey instrument[8]

1. You are invited to participate in a research study about American adults' perceptions of political messages on Facebook and Twitter. Your participation is voluntary. In order to make this study a valid one, you will not be made fully aware of the exact purpose of the study. By participating, you are agreeing to that alteration.

In this study, you will be asked to do the following things:

- Login to MTurk as an ap-proved participant
- Take a survey on MTurk about political messages on social media

If you have any questions, concerns, or complaints about this study, now or in the future, please contact Dr. Jason Ross Arnold at jrarnold@vcu.edu.

- o I consent
- o I do not consent

We would like to begin by asking you a few questions about yourself.

2. In what year were you born? _____ (enter 4-digit birth year; for example, 1976)

3. When it comes to politics, do you usually think of yourself as ...

- o Extremely liberal
- o Liberal
- o Slightly liberal
- o Moderate, middle of the road
- o Slightly conservative
- o Conservative
- o Extremely conservative
- o Haven't though much about it

4. Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or something else?

- o Republican
- o Democrat
- o Independent
- o Something else

5. If you think of yourself as a Republican, would you call yourself a strong Republican or a not very strong Republican?

- o Strong Republican
- o Not very strong Republican

---

[8] Branching logic is embedded in SurveyMonkey, and not shown in this printed questionnaire.

6. If you think of yourself as a Democrat, would you call yourself a strong Democrat or a not very strong Democrat?
- o   Strong Democrat
- o   Not very strong Democrat

7. If you think of yourself as an Independent or something else, do you think of yourself as closer to the Republican Party or the Democratic Party?
- o   The Republican Party
- o   The Democratic Party

Now we would like to ask you some questions about your media use habits.

8. Which social media platform do you use the more often?
- o   Facebook
- o   Twitter
- o   I don't use Facebook or Twitter

9. How frequently do you use social media?
- o   Several times an hour
- o   Several times a day
- o   Once a day
- o   Several times a week
- o   Once a week
- o   Less often than once a week
- o   Never

10. How much time, on average, do you spend on social media on the days that you use it?
- o   Less than an hour
- o   1-2 hours
- o   3-5 hours
- o   More than five hours

11. When you use social media, how likely are you to share political or issue- based posts (by reposting, liking, or commenting)?
- o   Very unlikely
- o   Unlikely
- o   Likely
- o   Very likely

12. When you use social media, how likely are you to share non-political posts, such as about every day or special personal experiences (by reposting, liking, or commenting)?
- o   Very unlikely
- o   Unlikely
- o   Likely
- o   Very likely

13. Is the time that you spend on social media for personal or business use?
- o Business
- o Personal
- o Both

14. Although you don't use Facebook or Twitter personally, which social media platform are you more familiar with?
- o Facebook
- o Twitter

Now we would like to ask you some questions about your media use habits.

15. Which social media platform do you use the more often?
- o Facebook
- o Twitter
- o I don't use Facebook or Twitter

16. How frequently do you use social media?
- o Several times an hour   Several times a day   Once a day
- o Several times a week   Once a week
- o Less often than once a week   Never

17. How much time, on average, do you spend on social media on the days that you use it?
- o Less than an hour
- o 1-2 hours
- o 3-5 hours
- o More than five hours

18. When you use social media, how likely are you to share political or issue- based posts (by reposting, liking, or commenting)?
- o Very unlikely
- o Unlikely
- o Likely
- o Very likely

19. When you use social media, how likely are you to share non-political posts, such as about every day or special personal experiences (by reposting, liking, or commenting)?
- o Very unlikely
- o Unlikely
- o Likely
- o Very likely

20. Is the time that you spend on social media for personal or business use?
- o Business
- o Personal
- o Both

21. Although you don't use Facebook or Twitter personally, which social media platform are you more familiar with?
   o   Facebook
   o   Twitter

Now we're going to show you an example from Facebook, followed by questions about the Facebook post.

Please take a few moments to read the post below.

It is important that you have a good idea of the information being presented in this post because you will be asked questions about it when you move to the next page.

[Images from Figures 3 or 4 presented to subjects]

You will be asked questions about this post when you move to the next page.

22. Are you ready to proceed?
   o   Yes
   o   No

Now we're going to show you an example from Twitter, followed by questions about the tweet.

Please take a few moments to read the tweet below.

[Images from Figures 3 or 4 presented to subjects]

It is important that you have a good idea of the information being presented in this tweet because you will be asked questions about it when you move to the next page.

You will be asked questions about this tweet when you move to the next page.

23. Are you ready to proceed?
   o   Yes
   o   No

Now we're going to show you an example from Facebook, followed by questions about the Facebook post.

Please take a few moments to read the post below.

[Images from Figures 3 or 4 presented to subjects]

It is important that you have a good idea of the information being presented in this post because you will be asked questions about it when you move to the next page.

24. You will be asked questions about this post when you move to the next page. Are you ready to proceed?
- o Yes
- o No

Now we're going to show you an example from Twitter, followed by questions about the tweet.

Please take a few moments to read the tweet below.

[Images from Figures 3 or 4 presented to subjects]

It is important that you have a good idea of the information being presented in this tweet because you will be asked questions about it when you move to the next page.

25. You will be asked questions about this tweet when you move to the next page. Are you ready to proceed?
- o Yes
- o No

26. Imagine that you came across this post on your social media feed. How likely is it that you would "like" this post on social media?
- o Very likely
- o Likely
- o Somewhat likely
- o Neither likely nor unlikely
- o Somewhat unlikely
- o Unlikely
- o Very unlikely

27. Imagine that you came across this post on your social media feed. How likely is it that you would "share" this post on social media?
- o Very likely
- o Likely
- o Somewhat likely
- o Neither likely nor unlikely
- o Somewhat unlikely
- o Unlikely
- o Very unlikely

28. Imagine that you came across this post on your social media feed.
How likely is it that you would initiate a conversation about the information in this post with a friend, family member, acquaintance, or anyone else?
- o Very likely
- o Likely
- o Somewhat likely
- o Neither likely nor unlikely
- o Somewhat unlikely
- o Unlikely
- o Very unlikely

29. Imagine that you came across this post on your social media feed. How likely is it that you would talk about the information in this post in a conversation with a friend or family member?
    o Very likely
    o Likely
    o Somewhat likely
    o Neither likely nor unlikely
    o Somewhat unlikely
    o Unlikely
    o Very unlikely

30. Thinking about the statement in the post, would you consider this claim to be true, somewhat true, somewhat false, or false?
    o True
    o Somewhat true
    o Somewhat false
    o False

31. Thinking about the content of the post, did you find the post very useful, somewhat useful, not useful at all?
    o Very useful
    o Somewhat useful
    o Not useful at all
    o Don't know

32. Thinking about the tone of the post, would you characterize the tone of the post as overly hostile, somewhat hostile, or not hostile at all?
    o Overly hostile
    o Somewhat hostile
    o Not hostile at all
    o Don't know

33. Some states allow voters to cast their ballots early or to mail in their ballots. Have you already voted in the November 2020 election?
    o Yes
    o No

34. Does the information in this post make you more likely to vote in the upcoming November 2020 election, or less likely to vote?
    o Much more likely
    o Somewhat more likely
    o Neither more or less likely
    o Somewhat less likely
    o Much less likely

35. There are many ways to cast your ballot in the upcoming November 2020 election. In some states you can mail in your vote. If your state allowed mail-in voting would the information in this post make you more likely to vote by mail in the 2020 election, or less likely to vote by mail?
- o   Much more likely
- o   Somewhat more likely
- o   Neither more or less likely
- o   Somewhat less likely
- o   Much less likely

36. If you hadn't yet cast your ballot, and if your state allowed mail-in voting, do you think the information in this post would have made you more likely or less likely to vote by mail?
- o   Much more likely
- o   Somewhat more likely
- o   Neither more or less likely
- o   Somewhat less likely
- o   Much less likely

37. How likely is it that you would participate in a demonstration or protest in your area based on the information in this post?
- o   Very likely
- o   Likely
- o   Somewhat likely
- o   Neither likely nor unlikely
- o   Somewhat unlikely
- o   Unlikely
- o   Very unlikely

38. Imagine that you came across this tweet on your social media feed. How likely is it that you would "like" this tweet on social media?
- o   Very likely
- o   Likely
- o   Somewhat likely
- o   Neither likely nor unlikely
- o   Somewhat unlikely
- o   Unlikely
- o   Very unlikely

39. Imagine that you came across this tweet on your social media feed. How likely is it that you would "share" this tweet on social media?
- o   Very likely
- o   Likely
- o   Somewhat likely
- o   Neither likely nor unlikely
- o   Somewhat unlikely
- o   Unlikely
- o   Very unlikely

40. Imagine that you came across this tweet on your social media feed.
How likely is it that you would initiate a conversation about the information in this tweet with a friend, family member, acquaintance, or anyone else?
- o   Very likely
- o   Likely
- o   Somewhat likely
- o   Neither likely nor unlikely
- o   Somewhat unlikely
- o   Unlikely
- o   Very unlikely

41. Imagine that you came across this tweet on your social media feed. How likely is it that you would talk about the information in this tweet in a conversation with a friend or family member?
- o   Very likely
- o   Likely
- o   Somewhat likely
- o   Neither likely nor unlikely
- o   Somewhat unlikely
- o   Unlikely
- o   Very unlikely

42. Thinking about the statement in the tweet, would you consider this claim to be true, somewhat true, somewhat false, or false?
- o   True
- o   Somewhat true
- o   Somewhat false
- o   False

43. Thinking about the content of the tweet, did you find the post very useful, somewhat useful, not useful at all?
- o   Very useful
- o   Somewhat useful
- o   Not useful at all
- o   Don't know

44. Thinking about the tone of the tweet, would you characterize the tone of the tweet as overly hostile, somewhat hostile, or not hostile at all?
- o   Overly hostile
- o   Somewhat hostile
- o   Not hostile at all
- o   Don't know

45. Some states allow voters to cast their ballots early or to mail in their ballots. Have you already voted in the November 2020 election?
- o   Yes
- o   No

46. Does the information in this tweet make you more likely to vote in the upcoming November 2020 election, or less likely to vote?
   o   Much more likely
   o   Somewhat more likely
   o   Neither more or less likely
   o   Somewhat less likely
   o   Much less likely

47. There are many ways to cast your ballot in the upcoming November 2020 election. In some states you can mail in your vote.
If your state allowed mail-in voting would the information in this tweet make you more likely to vote by mail in the 2020 election, or less likely to vote by mail?
   o   Much more likely
   o   Somewhat more likely
   o   Neither more or less likely
   o   Somewhat less likely
   o   Much less likely

48. If you hadn't yet cast your ballot, and if your state allowed mail-in voting, do you think the information in this tweet would have made you more likely or less likely to vote by mail?
   o   Much more likely
   o   Somewhat more likely
   o   Neither more or less likely
   o   Somewhat less likely
   o   Much less likely

49. How likely is it that you would participate in a demonstration or protest in your area based on the information in this tweet?
   o   Very likely
   o   Likely
   o   Somewhat likely
   o   Neither likely nor unlikely
   o   Somewhat unlikely
   o   Unlikely
   o   Very unlikely

50. Do you agree or disagree with the following statement:
Americans can trust our electoral system to produce a legitimate winner in the presidential election.
   o   Strongly agree
   o   Agree
   o   Slightly agree
   o   Neither agree nor disagree
   o   Slightly disagree
   o   Disagree
   o   Strongly disagree

51. Do you agree or disagree with the following statement:
I will participate in a demonstration or protest if I think the outcome of the presidential election is rigged.
- o Strongly agree
- o Agree
- o Slightly agree
- o Neither agree nor disagree
- o Slightly disagree
- o Disagree
- o Strongly disagree

52. In general how concerned are you that our elections can be rigged?
- o Very concerned
- o Somewhat concerned
- o Not very concerned
- o Not concerned at all

The next questions are about all other kinds of media, as well as your opinions about the news.

53. How often do you watch, read, or hear news? Do you generally watch, read, or hear news...
- o Several times an hour
- o Several times a day
- o Once a day
- o Several times a week
- o Once a week
- o Less often than once a week
- o Never

54. How important is it to you personally to keep up with news and information?
- o Not at all important
- o Not very important
- o Somewhat important
- o Very important
- o Extremely important

55. How do you get most of your news about national and international issues? Please rank order these news sources from 1 (you use this outlet least often for news) to 8 (you use this outlet most often for news):
- o Local TV stations (e.g., Chicago's CBS station)
- o National TV network stations (national networks, such as ABC, CBS, NBC)
- o Cable TV stations (e.g., CNN, Fox News, MSNBC)
- o Radio stations (AM or FM)
- o News websites (e.g., CNN.com, New York Times at https:/ www.nytimes.com/)
- o Blogs, Medium.com, etc.
- o Printed newspapers, magazines (e.g., print version of New York Times)
- o Social media

56. Choose the statement that best describes you, even if it is not exactly right. In general,
   o   I actively seek out news and information.
   o   I mostly bump into news and information as I do other things or hear about it from others.

57. Some people don't pay much attention to politics generally. From day to day, when there isn't any big election campaign going on, would you say you are...
   o   Not interested at all
   o   Not very interested
   o   Somewhat interested
   o   Very interested
   o   Extremely interested

58. To what extent do you agree or disagree with the following statement: I feel I have a pretty good understanding of the important political issues facing America.
   o   Strongly agree
   o   Agree
   o   Neither agree nor disagree
   o   Disagree
   o   Strongly disagree

59. Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?
   o   Most people can be trusted
   o   You can't be too careful in dealing with people

60. How much of the time do you think you can trust the government in Washington, DC to do what is right?
   o   Just about always
   o   Most of the time
   o   Only some of the time
   o   Almost never

61. To what extent do you think the mass media – such as newspapers, TV, and radio – reports the news accurately?
   o   A great deal
   o   A moderate amount
   o   Not much
   o   Not at all

62. To what extent do you agree with this statement: Twitter and Facebook can police the misuse of the social media platform.
   o   Strongly disagree
   o   Disagree
   o   Somewhat disagree
   o   Neither agree nor disagree
   o   Somewhat agree
   o   Agree
   o   Strongly agree

Next, we would like to ask you a few questions about your views about child rearing. Although there are a number of qualities that people feel that children should have, every person thinks that some are more important than others. Please indicate which trait is more important for a child to have.

63. Which one is more important for a child to have?
   o   Independence
   o   Respect for Elders

64. Which one is more important for a child to have?
   o   Curiosity
   o   Good Manners

65. Which one is more important for a child to have?
   o   Obedience
   o   Self-Reliance

66. Which one is more important for a child to have?
   o   Being Considerate
   o   Well Behaved

67. The next set of questions helps us learn what types of information are commonly known to the public. Many people don't know the answers to these questions, but we'd be grateful if you would please answer every question even if you're not sure what the right answer is.

It is important to us that you do NOT use outside sources like the Internet to search for the correct answer. If you answer No to this question you will be disqualified -- and, therefore, not paid -- because you are telling us that you intend to look up outside information even though we've asked you not to. Will you answer the following questions without help from outside sources?
   o   Yes
   o   No

68. Do you happen to know what job or political office is now held by Mike Pence?
   o   Don't know
   o   Yes (please specify)

69. Whose responsibility is it to determine if a law is constitutional or not? Is it…?
   o   The president
   o   The Congress
   o   The Supreme Court
   o   Don't Know

70. How much of a majority is required for the U.S. Senate and House to override a presidential veto?
   o   Don't know
   o   Yes (please specify)

71. Do you happen to know which party has the most members in the House of Representatives in Washington, DC?
   o   Don't know
   o   Yes (please specify)

72. Would you say that one of the parties is more conservative than the other at the national level?
   o   Don't know
   o   No
   o   Yes - Which party is more conservative (please specify)

We would like to end by asking you some basic questions about your background and identity.

73. What state do you reside in?
   o   [DROP DOWN MENU WITH ALL STATES]

74. What is the highest level of education you have completed?
   o   Did not attend school
   o   1st grade
   o   2nd grade
   o   3rd grade
   o   4th grade
   o   5th grade
   o   6th grade
   o   7th grade
   o   8th grade
   o   9th grade
   o   10th grade
   o   11th grade
   o   Graduated from high school
   o   1 year of college
   o   2 years of college
   o   3 years of college
   o   Graduated from college
   o   Some graduate school
   o   Completed graduate school

75. What is your gender?
   o   Female
   o   Male

76. How do you describe yourself?
   o   Male
   o   Female
   o   Transgender
   o   Do not identify as male, female, or transgender
   o   Prefer not to say

77. Are you of Spanish, Hispanic, or Latino descent?
   o   Yes
   o   No
   o   Prefer not to say

78. Since you've selected Spanish, Hispanic, or Latino in the previous question, we would like to find out more about your background. Please tell us which of the following categories do you identify yourself as.
   o   Mexican, Mexican-American, or Latino
   o   Puerto Rican
   o   Cuban
   o   Central American
   o   South American
   o   Caribbean
   o   Other Spanish/Hispanic/Latino   Prefer not to say

79. Which of the following categories indicates which race or ethnicity you consider yourself to be? You can check more than one box.
   o   White
   o   Black or African American
   o   American Indian or Alaskan Native
   o   Asian Indian
   o   Chinese
   o   Filipino
   o   Japanese
   o   Korean
   o   Vietnamese
   o   Other Asian
   o   Native Hawaiian
   o   Guamanian or Chamorro
   o   Samoan
   o   Other Pacific Islander
   o   Prefer not to say
   o   Some other race (please specify)

80. It is essential for the validity of this study that we know whether participants looked up any information online during the study. Did you make an effort to look up information during the study? Please be honest; you will not be penalized in any way if you did.
   o   Yes, I looked up information
   o   No, I did not look up information

81. Do you have any comments on the survey? Please let us know about any problems you had or aspects of the survey that were confusing.

Please read through the following information about the research study you just participated in.

At the very bottom you will see your validation code.
Thank you for your participation in this experimental research study.

Through this study we hope to gain a better idea of the effects of social media on misinformation, specifically as it pertains to foreign interference campaigns. In particular, the purpose of the experiment is to evaluate the how Twitter and Facebook source cues influence the acceptance of the Twitter or Facebook message and the likelihood that an individual will chose to retweet, share, or otherwise disseminate the message. The study is also interested in how misinformation on Twitter and Facebook can impact social media users' opinions about the legitimacy of the upcoming presidential election and their likelihood of voting and participating in election-related protests. In order to better understand these effects, you, like all other participants, were assigned to a random treatment group.

The Twitter or Facebook posts you may have been exposed to included intentionally misleading information about mail-in voting. These were designed to reflect real memes that promote this sort of misinformation. Please be advised that you should not use this misinformation when making evaluations about politics.

For more information on why the Wisconsin meme was factually inaccurate, please visit this site: https://www.usatoday.com/story/news/factcheck/2020/08/17/fact-check-wisconsin-mailbox-photo-not-proof-voter-suppression/3382701001.

For more information on why the Michigan meme was factually inaccurate, please visit this site: https:/apnews.com/article/fact-checking-9245320034.

Remember, this is an anonymous survey. The results may be published, but your name will not be. Please do not hesitate to contact the study's principal investigator, Dr. Jason Arnold, if you have questions about these studies. His email address is jrarnold@vcu.edu.

Again, thank you for your participation.

Enter this validation code into Mechanical Turk to complete your HIT: imQ2h1w3QnU08sBJvwJb
Enter this validation code into Mechanical Turk to complete your HIT: eD53kJ2uUKOn4MLrd9Hh
Enter this validation code into Mechanical Turk to complete your HIT: PrYx7kMZLbYP8IP1iEuN
Enter this validation code into Mechanical Turk to complete your HIT: sn9h5vsiPB1dPeU9Rg6v
Enter this validation code into Mechanical Turk to complete your HIT: 2HaQrOBBL6O0xLoJ2uOj
Enter this validation code into Mechanical Turk to complete your HIT: JWO6M4gdVM9KMkistuNQ
Enter this validation code into Mechanical Turk to complete your HIT: qL2kwaJ5AFZCcHBSqJzs
Enter this validation code into Mechanical Turk to complete your HIT: insrRIRTro06OR243AGo
Enter this validation code into Mechanical Turk to complete your HIT: YpFB3y1nHnPaJt65LDq5
Enter this validation code into Mechanical Turk to complete your HIT: a5cGLQjkrAM5yvoAxbt3
Enter this validation code into Mechanical Turk to complete your HIT: QsGBu62keuuC7U35JT75
Enter this validation code into Mechanical Turk to complete your HIT: 7CT4h4TJxTzISNzZRMO4
Enter this validation code into Mechanical Turk to complete your HIT: MDwcRLrgzMnvOT6Pe4RK
Enter this validation code into Mechanical Turk to complete your HIT: pgypNLZs5vjU0gvrNB55
Enter this validation code into Mechanical Turk to complete your HIT: XW3WbkHOe7P81OLwZGMX
Enter this validation code into Mechanical Turk to complete your HIT: DBKMmdG5Gzbfh4n308cD
Enter this validation code into Mechanical Turk to complete your HIT: E821iz99ruGaSnfG8Yno
Enter this validation code into Mechanical Turk to complete your HIT: NOLDrw8AZVZCtWjxWNe5
Enter this validation code into Mechanical Turk to complete your HIT: eycHs60Qyh6XmkLLNijV
Enter this validation code into Mechanical Turk to complete your HIT: xGwtJJVHI0qrKchWLcuJ

Appendix C: Independent variable descriptions

Appendix Table B. *Means and standard deviations of independent variables.*

	Facebook Republicans	Facebook Democrats	Twitter Republicans	Twitter Democrats
Age	39.08 (10.97)	36.90 (10.64)	39.37 (11.65)	34.61 (10.84)
Education	3.79 (1.16)	3.93 (1.19)	3.74 (1.28)	3.80 (1.13)
Female	43.2%	44.6%	29.1%	35.4%
South	39.7%	39%	43%	35%
Authoritarianism	2.09 (1.29)	1.63 (1.43)	2.02 (1.39)	1.35 (1.33)
Share Political Information	2.72 (0.95)	2.55 (0.89)	2.54 (0.95)	2.50 (0.90)
Social Media Use	4.83 (1.06)	4.62 (1.09)	4.60 (1.15)	4.77 (0.93)

Note: Means (Standard Deviation). Female and South are percent Female and percent living in the South.

- **Age** is a continuous variable ranging from 19-79.

- **Education** is categorical. 0 = no high school, 1 = some high school, 2 = high school, 3 = some college, 4 = college grad, 5 = some graduate school, 6 = graduate school grad.

- **Share political information** is coded 1-4, where 1 = very unlikely, 2 = unlikely, 3 = likely, 4 = very likely.

- **Social media use** is coded 0 to 6, where 0 = never, 1 = less often than once a week, 2 = once a week, 3 = several times a week, 4 = once a day, 5 = several times a day, 6 = several times an hour

- **Gender** is coded so that Male = 0, Female = 1.

- **South** is coded so that other regions = 0, South = 1. Region codes based on U.S. Census (https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf).

- **Authoritarianism** is an index from 0-4 and is based on Feldman and Sterners (1997) child rearing questions. Specifically, subjects were asked four questions: "Which one is more important for a child to have? (Independence/Respect for Elders, Curiosity/Good Manners, Obedience/Self-Reliance, Being Considerate/Well Behaved)." Subjects who indicated "Independence," "Curiosity," "Self-Reliance," and "Being Considerate" scored a minimum on the scale of zero; while those who indicated "Respect for Elders," "Good Manners," "Obedience," and "Well Behaved" scored a maximum of 4.