Going Public in an Era of Social Media: Tweets, Corrections, and Public Opinion

Dino P. Christenson
Washington University in St. Louis
Sarah Kreps
Cornell University
Douglas L. Kriner
Cornell University

Forthcoming 2021 in *Presidential Studies Quarterly*

Abstract

Presidents invariably use the bully pulpit to push a political agenda, but whether this leads to political success in advancing that agenda has long been the subject of debate. The increased reliance on social media has renewed that debate, particularly in light of new policies that flag or remove objectionable presidential content. This research conducts a survey experiment that evaluates the effect of presidential tweets on support for executive policies, including proposed unilateral action, and studies the effect of social media corrections of those tweets. We find little evidence that social media appeals move public opinion overall, although they do increase support among Republicans. Corrections generally worked as intended among Democrats but backfired among Republicans, cancelling each other out in the aggregate. The findings offer important insights into the efficacy of going public on social media and of corrections to such claims in an era of stark partisan polarization.

On Jan 20, 2017, President-Elect Trump took to Twitter: "It all begins today! I will see you at 11:00 A.M. for the swearing-in. THE MOVEMENT CONTINUES – THE WORK BEGINS!" The tweet foreshadowed his reliance on the platform for announcing new policies, attacking adversaries, and bolstering support among his base. Within the first 33 months of Trump's presidency, he tweeted more than 11,000 times (Shear et al. 2019). While the veracity of those tweets long attracted scrutiny (Oiu 2020), for years Twitter allowed unfounded, conspiratorial, or misleading presidential tweets to go unscathed.

In May 2020, however, social media platforms began actively fact-checking presidential claims (Conger and Isaac 2020) and even removing presidential content that violates corporate content policies (Glazer 2020). Past research on corrections has yielded decidedly mixed evidence of their efficacy (Clayton et al. 2019; Thorson 2016; Wood and Porter 2019). Are corrections effective when they seek to rebut false claims when the source is the president and not a Russian bot or random social media post? Do they counteract any efficacy of the original miscommunication, or even reduce false perceptions below a baseline level? Or might corrections trigger a backfire effect, at least among those predisposed to believe the president's claim, and actually reinforce misperceptions (Nyhan and Reifler 2010)?

We answer these questions with a study of presidential tweets and corrections. As the subject of our analysis, we use an episode that incorporated debates about freedom of speech and executive authority. Within days of Twitter first flagging one of President Trump's tweets as false, President Trump issued a retaliatory executive order targeting the legal protections enjoyed by social media platforms. Through an original survey experiment that probes whether exposure to Trump's tweet and corrections to it affect public support for the executive action, we find little evidence that social media appeals move public opinion; the effects we do observe are limited to

his co-partisans. Similarly, corrections along the lines that social media firms have begun issuing in response to presidential tweets, and even more aggressive corrections, did little to reduce misperceptions. However, the null effects for corrections in the aggregate mask important differences in partisan responses. Among those of the opposition party, corrections generally worked as intended whereas among co-partisans, they generally backfired, suggesting that corrections triggered significant partisan motivated reasoning.

Our research makes direct contributions to three different literatures at the intersection of presidency scholarship and political communication. First, we engage the venerable literature on presidential public appeals (Cohen 2010; Edwards 2006; Kernell 1997) by examining their efficacy in the social media age. In so doing, we build on an emerging literature that uses survey experiments to measure the efficacy of going public (e.g. Cohen 2015), with a specific focus here on the efficacy of a prominent presidential tweet. In alleging that mail-in ballots inevitably produce systemic electoral fraud, President Trump tried both to grow and reinforce public doubts about the integrity of the electoral system and the extent of fraud, and to rally public opinion against legislative proposals such as the Natural Disaster and Emergency Ballot Act of 2020, which would dramatically expand access to absentee ballots during the coronavirus pandemic.

Second, our results contribute to the literature on corrections to false information and their efficacy (Nyhan and Reifler 2010; Clayton et al. 2019; Thorson 2016; Wood and Porter 2019). Importantly, we examine the efficacy of corrections through a new lens in an intensely polarized setting where the false claim is advanced directly by the president. Finally, we also contribute to the nascent literature on public support for unilateral action (Christenson and Kriner 2017a; Lowande and Gray 2017; Reeves and Rogowski 2016) by examining whether support for

a prominent executive order is responsive to political rhetoric and social media corrections to such claims.

Going Public in the Social Media Age

Since Teddy Roosevelt declared the presidency a "bully pulpit," scholars and politicians alike have examined and debated the efficacy of presidential public appeals. Kernell's (1997) seminal work, *Going Public*, observed that presidents routinely appeal for public support for the administration's agenda. That presidents make public appeals does not mean that the strategy is effective. Indeed, empirical evidence that presidential appeals can move public opinion is decidedly mixed (e.g. Canes-Wrone 2005; Cavari 2017; Edwards 2006; Rottinghaus 2010). One camp shows that despite being the focal point of American politics and enjoying a preeminent position in shaping media coverage (Entman 2004), presidents routinely struggle to move public opinion and increase support for their policy positions and initiatives (Edwards 2006, 2009, 2016; Franco, Grimmer, and Lim 2018). Some studies have even raised fears that presidential appeals can backfire and effect shifts in public opinion away from the White House and its preferences (Cameron and Park 2011; Lee 2009).

A more optimistic camp has found greater evidence that presidential opinion leadership can be successful, but its influence is often highly conditional and varies across issues and groups (Canes-Wrone 2005; Cavari 2013, 2017; Cohen and Hamman 2003; Tedin, Rottinghaus, and Rodgers 2011; Wood 2007). One of the reasons that presidents struggle is that they must "break through the noise" of news coverage to truly lead public opinion (Eshbaugh-Soha and Peake 2011), but their ability to do so is conditional. Presidents are not the only actors who can go public (Grimmer 2013), and political contestation over messaging (Cameron and Park 2011),

which is bolstered by balance rules and the media's desire to index their coverage to the official debate in Washington (Bennett 1990), can blunt the force of presidential appeals.

Social media has opened a new channel for presidents to take their message directly to millions of Americans, circumscribing the mass media's gate-keeping role and balance norms that elevate competing voices and perspectives (Cameron and Park 2011). While political elites initially appeared to use social media primarily in a campaign context, driving the early research on social media as an arena for political communication (Christenson, Smidt, and Panagopoulos 2014; Enli and Skogerbø 2013; Jungherr 2016), they are increasingly using social media as a tool of governance. By offering presidents a direct conduit to millions of Americans, social media has the potential to circumvent a reliance on the mass media and bolster the efficacy of popular appeals. Trump intimated this point in a tweet: "Only the Fake News Media and Trump enemies want me to stop using Social Media (110 million people). Only way for me to get the truth out!"[1] Recent research appears to corroborate the intuition that social media offers a direct conduit to constituents, with evidence suggesting that Trump's intense social media engagement has had an agenda-setting effect (Lee and Xu 2018), as well as an influence on elite behavior (Fu and Howell 2020).

Whether and how targeted presidential appeals via social media influence public opinion remain open questions. One important exception examined the effect of President Trump's July 2018 tweet opposing public access to 3-D printed plastic guns. Given Trump's strong and vocal support for the Second Amendment, the tweet was "costly" rhetoric, which should increase the credibility of the signal (Baum and Groeling 2009; Calvert 1985). However, in a natural experiment Miles and Haider-Markel (2020) found little evidence that the tweet lowered support for 3-D gun printing. Indeed, the tweet appears to have backfired and actually increased support

for 3-D printed guns. Merely having millions of avid followers on Twitter appears to be no guarantee that a president can successfully exploit exposure to rally support for his positions.

Beyond the prevalence that characterizes Trump's social media use is the tendency to issue statements that independent factcheckers label false or misleading. An analysis of Trump's tweets found that about one-third in the period studied contained misleading information (Oiu 2020). For years, social media platforms resisted calls to flag or remove false claims. But on May 26, 2020, Twitter reversed its prior position and explicitly tagged one of the President's tweets alleging widespread electoral fraud from mail-in ballots as "potentially misleading." Beneath the tweet, Twitter added a hyperlink reading "Get the facts about mail-in ballots." Readers who clicked on the link were taken to a page that called Trump's claim "unsubstantiated," and that noted "experts say mail-in ballots are very rarely linked to fraud."

Twitter's move provoked a maelstrom of criticism from the White House and the president's partisan allies on Capitol Hill. Trump's retribution was swift and provocative, as the following day the President signed an executive order targeting social media companies and seeking to remove legal protections for moderating content that they enjoyed under Section 230 of the 1996 Communications Decency Act.

Previous research suggests that Trump need not have been concerned about Twitter's new policy of issuing corrections on unsubstantiated claims. While some scholars find that corrections can reduce the perceived accuracy of false claims (Clayton et al. 2019; Porter, Wood, and Kirby 2018; Wood and Porter 2019), others have found that the effects of corrections are modest (Garrett, Nisbet, and Lynch 2013; Thorson 2016) or highly conditional (Lewandowsky et al. 2012). Particularly in highly politicized contexts, such as misleading appeals made directly by the president, there is a potential for a backfire effect (Nyhan and Reifler 2010; Schwarz et al.

2007) in which corrections drive more traffic to a misinformation site, increase its fluency, and make corrections counterproductive (Berinsky 2017). However, other studies suggest that backfire effects are limited and uncommon in most contexts (Wood and Porter 2019). Although not examining appeals through social media per se, a recent study examining corrections to President Trump's false and misleading statements on climate change offers some insight (Porter, Wood, and Bahador 2019). Corrections did not produce a backfire effect and significantly reduced factual accuracy perceptions of Trump's claims; however, they had no effect on broader attitudes and policy preferences.

Research Design

To evaluate the effect of both going public and corrections issued on those public overtures, we conducted an original survey experiment fielded within days of President Trump's mail-in ballot fraud tweet. In general, experimental research has found stronger evidence for the efficacy of presidential appeals than studies analyzing primarily observational evidence (Cohen 2015; Druckman and Holmes 2004; Gillion 2017; Tedin, Rottinghaus, and Rodgers 2011). While experiments inevitably raise questions of external validity (Barabas and Jerit 2010), existing experimental evidence hints at the possibility that unmediated presidential communications may be more effective than more traditional speeches. As a test case, we examine the effect of exposure to President Trump's tweets about the existence and prevalence of voter fraud on public belief in these claims and preferences for the use of mail-in ballots in the 2020 election.

To shed new light on the efficacy of corrections, we examine the effect of three different corrections that vary by strength to Trump's statement on public beliefs and policy preferences.

Social media companies have openly acknowledged that one reason they have been reticent to take aggressive steps to counter misinformation is that they fear it could produce a backfire effect, driving increased traffic to fake news sites and actually bolstering popular beliefs in false claims. As a result, when they ultimately have decided to flag false claims, social media platforms have often employed nonspecific warnings that are designed more to nudge readers toward accurate information than to call out and rebut false claims (Porter and Wood 2020). However, more nuanced approaches risk undermining the efficacy of the correction by being too subtle (Avaaz 2020). We manipulate the nature of the correction—nudge versus explicit refutation—to examine whether the efficacy of a corrections varies with its strength, and also whether the strength of the correction affects the emergence and intensity of backfire effects among those predisposed to support President Trump and reject any correction to his claims.

Finally, we examine whether President Trump's tweet and the social media corrective response to it have any systematic effect on public support for the unilateral executive behavior that the social media correction provoked, specifically President Trump's executive order targeting legal protections enjoyed by social media companies. Past research argues that attitudes toward unilateralism are driven largely by partisan loyalties or policy preferences (Christenson and Kriner 2017a) or more general values, such as support for the rule of law (Reeves and Rogowski 2016). Moreover, the reaction of other political elites also affects public support, as institutional criticism can erode public support for executive action (Christenson and Kriner 2017b). Here, we examine the effect of pushback from a non-governmental actor that also takes a different form from direct criticism of the substance of a unilateral action. Specifically, we examine whether priming subjects to think about the motivation behind Trump's unilateral gambit – retaliation against Twitter for flagging his claims as unsubstantiated and

misleading – has any effect on public support for his executive order. Does it sap support, trigger a backlash, or have no effect? Answering this question allows us to examine the potential broader political consequences of aggressive corrections in an intensely polarized environment.

To test the efficacy of presidential social media appeals – and corrections to those appeals – in shaping popular attitudes and policy preferences, we embedded an experiment on an online survey fielded on June 9, 2020, just over a week after Trump's tweet alleging systematic fraud with mail-in ballots and social media corrections. The survey recruited a broadly representative sample of 1,003 adult Americans through the online marketplace Lucid. Lucid employs quota sampling to produce samples matched to the US population on age, gender, ethnicity, and geographic region (Coppock and McClellan 2019). The demographic composition of our sample and comparisons to those of prominent social science surveys and U.S. Census American Community Survey statistics are presented in SI Table 1.

After reading an informed consent form and agreeing to participate, subjects were randomly assigned to one of five experimental conditions. Those assigned to the control group received no information about Trump's tweet or any response by Twitter. Those in the first treatment group saw a graphic presenting President Trump's tweet from May 26: "There is no way (ZERO!) that Mail-in Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, ballots will be forced & fraudulently signed. The Governor of California is sending ballots to millions of people, anyone….." Comparing public opinion across the Trump treatment and control groups affords a conservative estimate of the efficacy of the tweet. The estimate is conservative as we cannot be sure how many subjects might have seen the tweet prior to the experiment. The tweet and Trump's response did attract considerable media attention. However, given well-documented deficiencies in most Americans' political knowledge (Delli

Carpini and Keeter 1997), we believe it likely that many were either unaware or paid only passing prior attention to the political dispute.

The final three experimental treatments examined the efficacy of corrections and how it might vary according to the scope of the correction. The treatments are displayed in Figure 1 below. In the first, subjects again saw the Trump tweet followed by the graphic/hyperlink used by Twitter to flag the tweet as misleading, which encouraged readers to click and "Get the facts about mail-in ballots." The second corrections treatment was identical to the first, but also included the information Twitter provided subjects if they clicked on the hyperlink about getting the facts. In this treatment, subjects also read a bolded headline "Trump makes unsubstantiated claim that mail-in ballots will lead to voter fraud." The headline was followed by some explanatory text labeling the claims "unsubstantiated" according to major media outlets, including CNN and the *Washington Post*. The correction concluded by stating that "Experts say mail-in ballots are very rarely linked to voter fraud." The final corrections treatment was modeled after Facebook corrections to misinformation about Covid-19. This correction labeled Trump's claim "factually incorrect" and provided direct evidence to rebut Trump's assertions. Specifically, it informed readers that "five states already use mail-in ballots and have reported little evidence of fraud." It also noted that experts agree that "there is little more than a handful of fraudulent mail-in ballots across the country each election," and it explained that the claims have been fact-checked by *Politifact* and others.[2]

[Figure 1 About Here]

All subjects were then asked the same two questions measuring beliefs about voter fraud in the United States. The first question directly asked about belief in Trump's charge: "Do you believe mail-in ballots lead to voter fraud?" Subjects could reply yes, no, or unsure. The second

question asked a slightly broader question about beliefs in fraud: "In general, how widespread do you think voter fraud is in U.S. elections? Do you think this happens a lot, sometimes, not much, or not at all?"

Later in the survey, subjects were also asked a pair of questions measuring their policy preferences. The first queried public support for expanded use of mail-in ballots in 2020 given the realities of Covid-19: "In response to the coronavirus pandemic, do you support or oppose allowing all U.S. citizens to vote by mail in the upcoming presidential election?" The second measured support for President Trump's executive order targeting social media companies. Subjects were asked, "last week, President Trump signed an executive order to limit the legal protections that federal law currently provides to social media platforms. Do you support or oppose this order?" Subjects answered both questions on a four-point Likert scale ranging from strongly support to strongly oppose.

Randomization checks showed no evidence of significant demographic imbalances across the five experimental conditions (see SI Table 2). As a result, to assess the effects of each treatment, we present the differences in mean opinion on each dimension across the relevant treatment group and the control.

Results

Figure 2 presents the aggregate effects of each experimental treatment on the beliefs and policy preferences of all subjects in the sample. For each treatment, dots illustrate the difference in mean opinion on the relevant dimension from that observed in the control group. I-bars present 95% confidence intervals about each difference in means.[3]

[Figure 2 About Here]

The top two panels of Figure 2 examine attitudes about voter fraud. The first most directly captures the essence of President Trump's core claim – that mail-in ballots lead to fraud. Forty-five percent of subjects believed that mail-in ballots do result in voter fraud in the control group. Exposure to President Trump's tweet had no effect as the percentage believing fraud in this condition was almost identical to that in the control. This null finding is consistent with research showing the uphill battle presidents face in changing public opinion, especially on issues where public opinion may have already calcified along partisan or ideological lines. Similarly, none of the corrections had any significant effects on beliefs in mail fraud. Even the strongest correction that presented subjects with the experience of states that already widely use mail-in ballots and the lack of evidence for systematic voting fraud failed to reduce beliefs in fraud. However, none of the corrections produced a backfire effect, at least in the aggregate.

The second panel of Figure 1 examines popular beliefs about the extent of voter fraud. In the control group, the mean on this measure was just under 3, which corresponds to beliefs that voter fraud occurs "sometimes." On this ordinal measure, exposure to the Trump tweet slightly increased perceptions about the prevalence of fraud from the control group baseline, while all three corrections treatments decreased it. None of the differences in means from the control were statistically significant. The difference in means across the Trump tweet and correction treatment was statistically significant ($p < .10$, two-tailed test) as was the difference in means between the Trump tweet condition and all three corrections treatments combined ($p < .10$, two-tailed test). However, the differences are substantively quite modest, not even accounting for a quarter point change on the four-point variable scale.

The third panel of Figure 2 examines support for expanded access to mail-in voting in 2020, the policy prescription that President Trump's tweet explicitly advocated against. Voting

by mail during the Covid-19 pandemic is very popular, with 68% support in the control group. Support for voting by mail was 3% lower in the Trump tweet treatment; however, the difference in means was not statistically significant. Similarly, none of the corrections that sought to combat misperceptions about vote-by-mail fraud had any effect in increasing support for mail-in voting.

The final panel examines support for President Trump's executive order to withdraw legal protections from social media platforms. Just over 50% of subjects supported the executive action in the control group. Support for the order was slightly higher in the Trump tweet condition, but the difference in means was not statistically significant. Support for the order was lower in all three of the corrections treatments than in the Trump tweet condition. However, none of the difference in means between any of the corrections groups and either the Trump tweet or control group baseline were statistically significant.

Effects by Party. In the aggregate, we found little evidence that either President Trump's tweet or efforts to correct it had a systematic effect on popular beliefs about voter fraud or on their policy preferences. This is consistent with literatures showing that presidential going public routinely fails to move the needle of public opinion, and that political misinformation can be stubbornly resistant to correction. However, it is also possible that the null effects in the aggregate mask significant variation in response to the treatments among different subsets of the public (e.g.Christenson and Glick 2015). Given prior research on partisan backfire effects (Cameron and Park 2011; Nyhan and Reifler 2010), it is possible that Democrats and Republicans responded to the same information in different ways, and that such swings could cancel each other out and yield no changes in opinion across treatments in the aggregate.

To examine this possibility and look for evidence of backfire effects, Figure 3 presents the effects of each treatment for Democrats and Republicans separately.[4] On the question of whether mail-in ballots produce voter fraud, we see a significant partisan gulf in the control condition with just over 30% of Democrats believing mail ballots cause fraud versus over 60% of Republicans. However, we continue to find little evidence that exposure to Trump's tweet had a significant effect on either partisan subgroup. Among Republicans the estimated effect was slightly positive; among Democrats it was slightly negative; but neither difference in means is statistically significant.

[Figure 3 About Here]

Corrections, by contrast, had dramatically different effects on the beliefs of Democrats and Republicans. Among Democrats, exposure to all three corrections of varying strength decreased belief in mail fraud; in the flagged and Twitter correction the decrease was 10%. Among Republicans, however, we observe the exact opposite. Belief in voter fraud was actually higher in all three corrections treatments than in the control group baseline – evidence of a backfire effect. In the correction treatment, which presented both the flag used by Twitter and its corrective information, belief that mail voter fraud occurs was more than 13% higher than in the control. While none of the partisan effects are significantly different from the control ($p < .05$, two-tailed test), in all four treatments the effects were of opposite directions for Democrats and Republicans, and in two conditions – the Twitter flag and correction treatments – the difference in partisan reactions exceeded 20%.[5] Thus, in the aggregate, there is little evidence that corrections were counterproductive. However, our partisan results suggest that social media companies' fears about backfire effects may be justified in highly politically charged contexts,

since these corrections increased misperceptions among those predisposed to believe President Trump.

We observe a similar pattern in beliefs about the prevalence of fraud more generally. All three corrections substantially reduced perceptions of the extent of voter fraud versus the control group baseline. Among Democrats, however, exposure to the Trump tweet had no effect. While among Republicans, exposure to President Trump's tweet increased beliefs about the prevalence of fraud, and none of the correction treatments significantly reduced fraud perceptions from the level observed in the Trump tweet treatment.[6]

Turning to support for voting by mail, we find little evidence that any of the treatments had an effect among Democrats. Almost 85% of Democrats supported greater access to absentee ballots in 2020 in the control group baseline. Trump's tweet did nothing to erode this support. And given a likely ceiling effect, corrections also failed to increase support further from its already very high base level. Just under 55% of Republicans in the control group supported voting by mail in 2020. Trump's tweet had no effect on this level of support. Similarly, exposure to corrections about Trump's false claims had no effect in rallying Republican support for mail voting during the pandemic. If anything, they may have somewhat decreased support, though none of the differences are statistically significant.

Finally, consistent with past research we find a very large partisan split in support for President Trump's executive order against social media companies. Fewer than 30% of Democrats backed the order in the control group, while more than 75% of Republicans did. Moreover, none of the treatments moved subjects off their partisan priors. Among Democrats, none of the treatments had any effect on support for the executive order. Among Republicans, exposure to Trump's tweet increased support for the order by 7%, though the difference in means

is not statistically significant. And again, corrections had no effect in eroding support for the executive order among Republicans. Support for the executive order across all three corrections treatments were statistically indistinguishable from the level of support observed in the Trump tweet treatment.

Conclusion

In the almost twenty-five years since the publication of Kernell's (1997) seminal work, an extensive literature has empirically examined the efficacy of presidential appeals in swaying public opinion and mobilizing popular support for the administration's agenda. Given the relative recency of Twitter as a vehicle for these appeals, questions of rhetorical efficacy and political persuasion remain unresolved. We take an important step in understanding whether presidential use of Twitter, a relatively new medium for political communication, moves public opinion, and whether the even more novel attempts at correcting misleading presidential posts on social media has any effect on Americans' accuracy perceptions of political claims and policy preferences.

Consistent with past scholarship warning that the bully pulpit does not guarantee presidents' influence over public opinion, we found little evidence that exposure to social media appeals promotes public convergence with the president's policy positions. In the aggregate, corrections to false presidential claims on social media also had little impact. However, this masks considerable heterogeneity across partisan groups. Exposure to various forms of corrections significantly affected beliefs about fraud, but by reducing misperceptions among those with partisan predispositions not to believe the president and increasing them among his co-partisans. Further, our analysis suggests that both social media appeals and corrections to

presidential claims had little influence on support for unilateral action to regulate social media platforms. Other factors, most importantly partisanship, anchored these assessments and rendered them unresponsive to claims and counter claims. Taken together, we find that the direct conduit between executives and the public—circumventing the media and communicating directly through an online platform—does not create unambiguously favorable effects for the president. Nor does correcting the misinformation universally arrive at its intended effects.

Our analysis was based on the first issue that prompted a change in social media correction policy, from one in which the president's tweets passed directly to his 80.3 million (and counting) followers (Lerman 2020) to one in which platforms such as Twitter flagged misleading content to users. Since Twitter labeled the initial election fraud tweet as misleading, it has also labeled executive tweets that warned of "serious force" for autonomous zones as "abusive behavior" (Feiner). Follow-on studies should broaden the analysis to include not just subsequent tweets by the US president to assess whether all issues operate similarly but also tweets by presidents of other countries to assess the cross-national dynamic. In March 2020, Twitter removed tweets by both the Brazilian and Venezuelan presidents for violating its coronavirus misinformation policies (Lyons 2020), providing both a political right and political left executive for comparison with the United States.

Finally, while we analyzed the near-term effects of social media appeals and corrections, we cannot rule out the potential for longer-term effects and the possibility that each misleading claim or correction has small effects that aggregate over time. Relatedly, some individual tweets may be outliers and they or their corrections may produce effects unobserved here. As the new approaches to content moderation on social media evolve and mature, we recommend a wider range of issues, additional presidents, and longitudinal effects as topics of further study.

Notes

[1] https://twitter.com/realdonaldtrump/status/892383242535481344.

[2] *Politifact* has ruled multiple claims by President Trump about mail-in ballots and voter fraud false. For example, see (Jacobson 2020).

[3] Logistic and ordered logit regressions with demographic controls yield substantively similar results. See SI Table 3.

[4] Subjects who "leaned" toward either party are coded as partisans. The results are robust to treating these subjects as Independents. See SI Figure 1.

[5] A logistic regression interacting each treatment with partisanship (SI Table 4) confirms that the effects of these two treatments on Republicans and Democrats were significantly different from one another.

[6] An ordered logit regression with partisan interactions (SI Table 4) confirms that the effect of each experimental treatment was significantly different across Republicans and Democrats.

References

Avaaz. 2020. *How Facebook Can Flatten the Curve of the Coronavirus Infodemic*.

Barabas, Jason, and Jennifer Jerit. 2010. "Are Survey Experiments Externally Valid?" *American Political Science Review* 104(2): 226–42.

Baum, Matthew A, and Tim Groeling. 2009. "Shot by the Messenger: Partisan Cues and Public Opinion Regarding National Security and War." *Political Behavior* 31(2): 157–86.

Bennett, W. Lance. 1990. "Toward a Theory of Press–State Relations in the United States." *Journal of Communication* 40(2): 103–27.

Berinsky, Adam J. 2017. "Rumors and Health Care Reform: Experiments in Political Misinformation." *British Journal of Political Science* 47(2): 241–62.

Calvert, Randall L. 1985. "The Value of Biased Information: A Rational Choice Model of Political Advice." *Journal of Politics* 47(2): 530–55.

Cameron, Charles, and Jee-Kwang Park. 2011. "Going Public When Opinion Is Contested: Evidence from Presidents' Campaigns for Supreme Court Nominees, 1930-2009." *Presidential Studies Quarterly* 41(3): 442–70.

Canes-Wrone, Brandice. 2005. *Who Leads Whom?: Presidents, Policy, and the Public*. Chicago: University of Chicago Press.

Cavari, Amnon. 2013. "The Short-Term Effect of Going Public." *Political Research Quarterly* 66(2): 336–51.

———. 2017. *The Party Politics of Presidential Rhetoric*. New York: Cambridge University Press.

Christenson, Dino P., and Douglas L. Kriner. 2017a. "Constitutional Qualms or Politics as Usual? The Factors Shaping Public Support for Unilateral Action." *American Journal of*

Political Science 61(2): 335–49.

———. 2017b. "Mobilizing the Public Against the President: Congress and the Political Costs of Unilateral Action." *American Journal of Political Science* 69(4): 769–85.

Christenson, Dino P., Corwin D. Smidt, and Costas Panagopoulos. 2014. "Deus Ex Machina: Candidate Web Presence and the Presidential Nomination Campaign." *Political Research Quarterly* 67(1): 108–22.

Christenson, Dino P, and David M Glick. 2015. "Chief Justice Roberts ' Health Care Decision Disrobed : The Microfoundations of the Supreme Court ' s Legitimacy ⇤." *American Journal of Political Science* 59(2): 403–18.

Clayton, Katherine et al. 2019. "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*: 1–23.

Cohen, Jeffrey E. 2010. *Going Local: Presidential Leadership in the Post-Broadcast Age*. New York: Cambridge University Press.

———. 2015. "Presidential Leadership of Public Opinion: An Embedded Survey Experiment." *Political Communication* 32(3): 345–55.

Cohen, Jeffrey E., and John A. Hamman. 2003. "The Polls: Can Presidential Rhetoric Affect the Public's Economic Perceptions?" *Presidential Studies Quarterly* 33(2): 408–22.

Conger, Kate, and Mike Isaac. 2020. "Defying Trump, Twitter Doubles Down on Labelling Tweets." *New York Times*.

Coppock, Alexander, and Oliver A. McClellan. 2019. "Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents." *Research and Politics* 6(1): 1–14.

Delli Carpini, Michael X, and Scott Keeter. 1997. *What Americans Know About Politics and Why It Matters*. New Haven, CT: Yale University Press.

Druckman, James N., and Justin W. Holmes. 2004. "Does Presidential Rhetoric Matter? Priming and Presidential Approval." *Presidential Studies Quarterly* 34(4): 755–78.

Edwards, George C. III. 2006. *On Deaf Ears: The Limits of the Bully Pulpit*. New Haven, CT: Yale University Press.

———. 2009. *The Strategic President: Persuasion and Opportunity in Presidential Leadership*. Princeton, NJ: Princeton University Press.

———. 2016. *Predicting the Presidency: The Potential for Persuasive Leadership*. Princeton, NJ: Princeton University Press.

Enli, Gunn Sara, and Eli Skogerbø. 2013. "PERSONALIZED CAMPAIGNS IN PARTY-CENTRED POLITICS: Twitter and Facebook as Arenas for Political Communication." *Information Communication and Society* 16(5): 757–74.

Entman, Robert M. 2004. *Projections of Power: Framing News, Public Opinion, and U.S. Foreign Policy.* Chicago: University of Chicago Press.

Eshbaugh-Soha, Matthew, and Jeffrey S. Peake. 2011. *Breaking Through the Noise: Presidential Leadership, Public Opinion, and the News Media*. Stanford, CA: Stanford University Press.

Feiner, Lauren. "Twitter Flagged Another Trump Tweet for Violating Its Policies." *CNBC*.

Franco, A, J Grimmer, and C Lim. 2018. "The Limited Effect of Presidential Public Appeals."

Fu, Shu, and William G. Howell. 2020. "The Behavioral Consequences of Public Appeals: Evidence on Campaign Fundraising from the 2018 Congressional Elections." *Presidential Studies Quarterly* 50(2): 325–47.

Garrett, R. Kelly, Erik C. Nisbet, and Emily K. Lynch. 2013. "Undermining the Corrective

Effects of Media-Based Political Fact Checking? The Role of Contextual Cues and Naïve Theory." *Journal of Communication* 63(4): 617–37.

Gillion, Daniel Q. 2017. "Obama's Discussion of Racial Policies and Citizens' Racial Resentment in the Experimental Setting." *Presidential Studies Quarterly* 47(3): 517–28.

Glazer, Emily. 2020. "Facebook Removes Trump Campaign Ads for Violating Policy on Use of Hate Symbol." *Wall Street Journal*.

Grimmer, Justin. 2013. *Representational Style in Congress: What Legislators Say and Why It Matters*. New York: Cambridge University Press.

Jacobson, Louis. 2020. "Donald Trump's Dubious Claim That 'Thousands' Are Conspiring on Mail-Ballot Fraud." *Politifact*.

Jungherr, Andreas. 2016. "Twitter Use in Election Campaigns: A Systematic Literature Review." *Journal of Information Technology and Politics* 13(1): 72–91.

Kernell, Samuel. 1997. *Going Public: New Strategies of Presidential Leadership*. Washington, DC: CQ Press.

Lee, Frances. 2009. *Beyond Ideology: Politics, Principles, and Partisanship in the U.S. Senate*. Chicago: University of Chicago Press.

Lee, Jayeon, and Weiai Xu. 2018. "The More Attacks, the More Retweets: Trump's and Clinton's Agenda Setting on Twitter." *Public Relations Review* 44(2): 201–13.

Lerman, Rachel. 2020. "Trump Says Twitter Is Trying to 'Silence' Conservatives. His Growing Number of Followers Suggests Otherwise." *Washington Post*.

Lewandowsky, Stephan et al. 2012. "Misinformation and Its Correction: Continued Influence and Successful Debiasing." *Psychological Science in the Public Interest, Supplement* 13(3): 106–31.

Lowande, Kenneth S., and Thomas Gray. 2017. “Public Perception of the Presidential Toolkit.” *Presidential Studies Quarterly* 47(3): 432–47.

Lyons, Kim. 2020. “Twitter Removes Tweets by Brazil, Venezuela Presidents for Violating Covid-19 Content Rules.” *The Verge*.

Miles, Matthew R., and Donald P. Haider-Markel. 2020. “Polls and Elections Trump, Twitter, and Public Dissuasion: A Natural Experiment in Presidential Rhetoric.” *Presidential Studies Quarterly* 50(2): 436–50.

Nyhan, Brendan, and Jason Reifler. 2010. “When Corrections Fail: The Persistence of Political Misperceptions.” *Political Behavior* 32(2): 303–30.

Oiu, Linda. 2020. “Hey @jack, Here Are More Questionable Tweets From @realdonaldtrump.” *New York Times*.

Porter, Ethan, and Thomas J. Wood. 2020. “Why Is Facebook So Afraid of Checking Facts?” *Wired*.

Porter, Ethan, Thomas J. Wood, and Babak Bahador. 2019. “Can Presidential Misinformation on Climate Change Be Corrected? Evidence from Internet and Phone Experiments.” *Research and Politics* 6(3): 205316801986478.

Porter, Ethan, Thomas J. Wood, and David Kirby. 2018. “Sex Trafficking, Russian Infiltration, Birth Certificates, and Pedophilia: A Survey Experiment Correcting Fake News.” *Journal of Experimental Political Science* 5(2): 159–64.

Reeves, Andrew, and Jon C. Rogowski. 2016. “Unilateral Powers, Public Opinion, and the Presidency.” *The Journal of Politics* 78(1): 137–51.

Rottinghaus, Brandon. 2010. *The Provisional Pulpit: Modern Presidential Leadership of Public Opinion*. College Station, TX: Texas A&M University Press.

Schwarz, Norbert, Lawrence J. Sanna, Ian Skurnik, and Carolyn Yoon. 2007. "Metacognitive Experiences and the Intricacies of Setting People Straight: Implications for Debiasing and Public Information Campaigns." *Advances in Experimental Social Psychology* 39: 127–61.

Shear, Michael et al. 2019. "How Trump Reshaped the Presidency in Over 11,000 Tweets." *New York Times*.

Tedin, Kent, Brandon Rottinghaus, and Harrell Rodgers. 2011. "When the President Goes Public: The Consequences of Communication Mode for Opinion Change across Issue Types and Groups." *Political Research Quarterly* 64(3): 506–19.

Thorson, Emily. 2016. "Belief Echoes: The Persistent Effects of Corrected Misinformation." *Political Communication* 33(3): 460–80.

Wood, B. Dan. 2007. *The Politics of Economic Leadership: The Causes and Consequences of Presidential Rhetoric*. Princeton, NJ: Princeton University Press.

Wood, Thomas, and Ethan Porter. 2019. "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence." *Political Behavior* 41(1): 135–63.

Figure 1: Strength of Correction Treatments

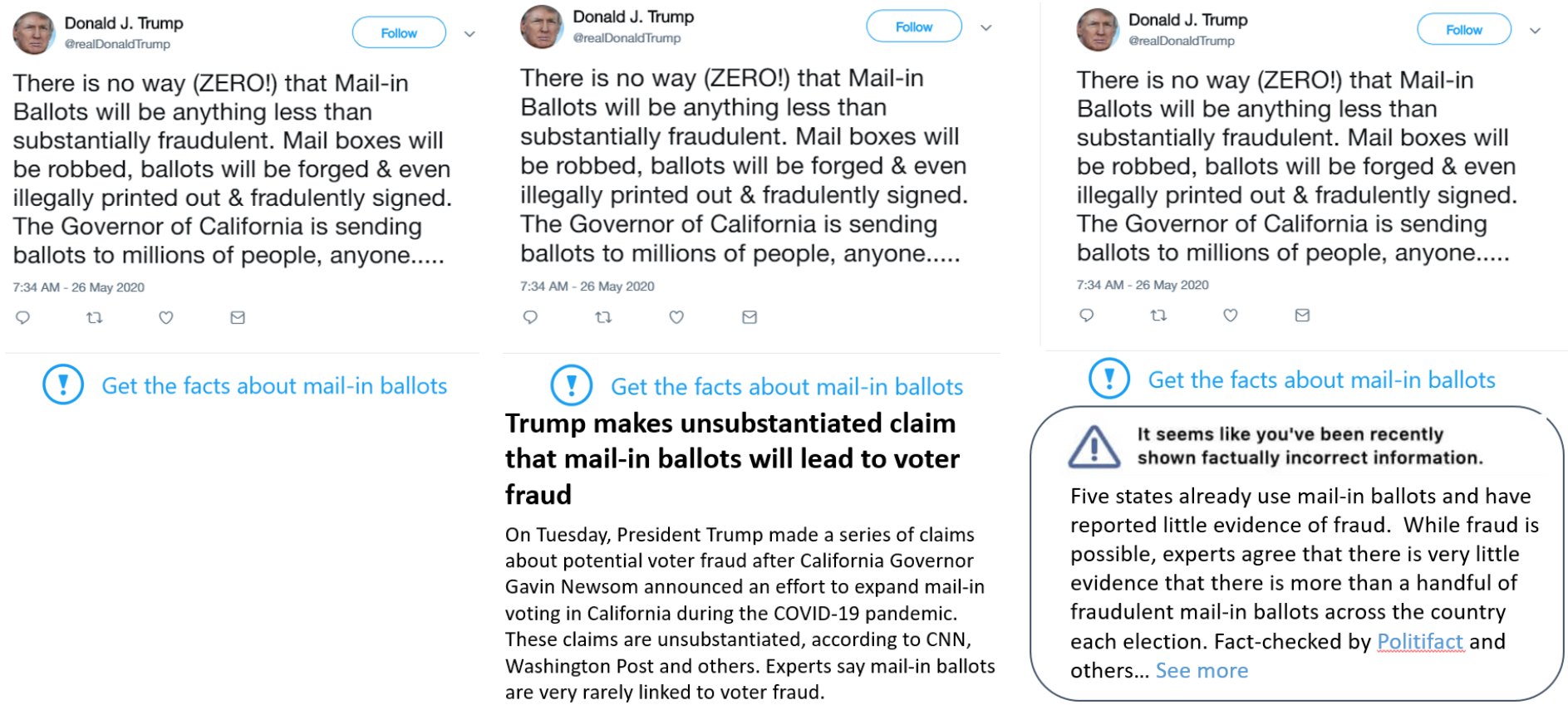
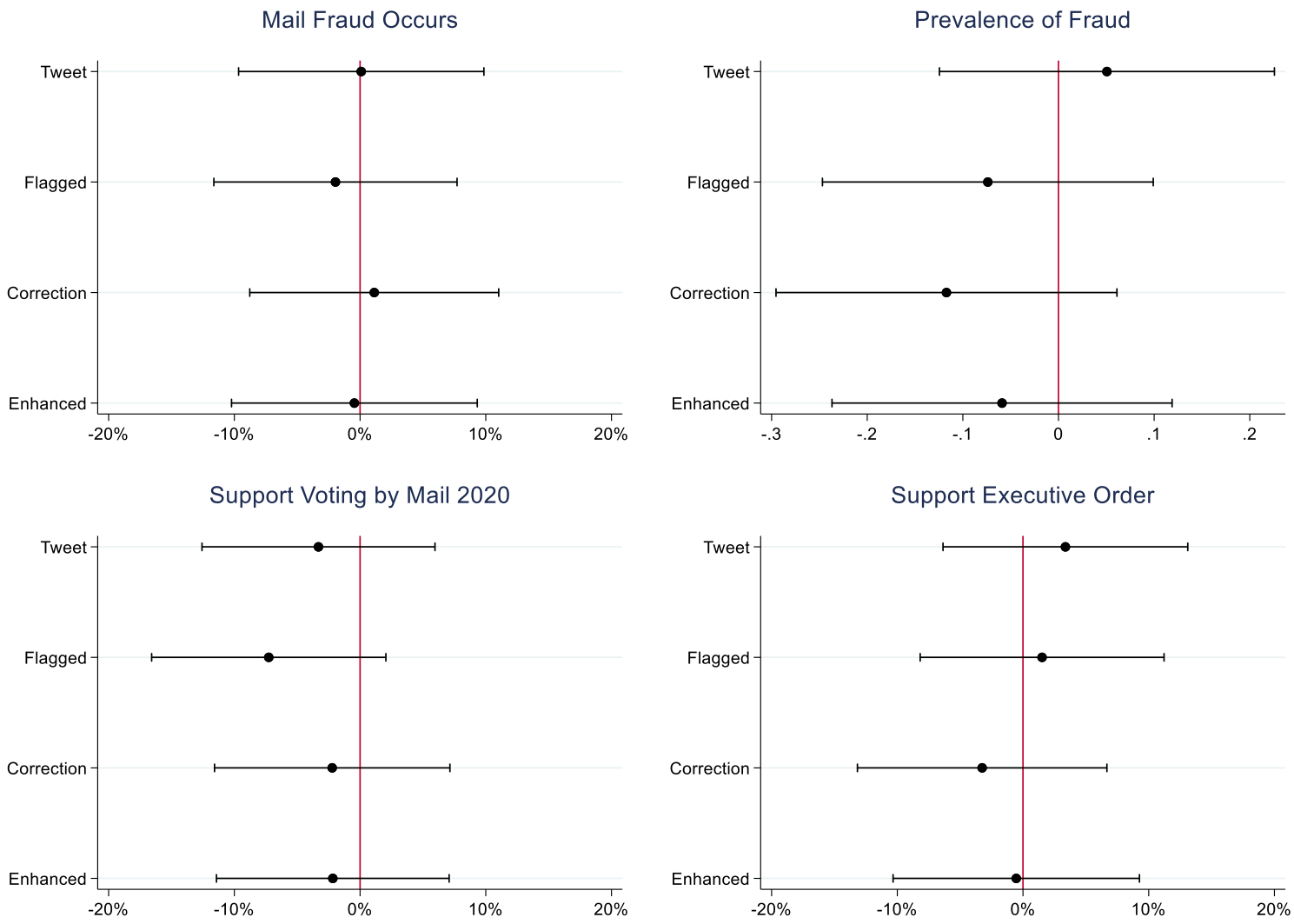
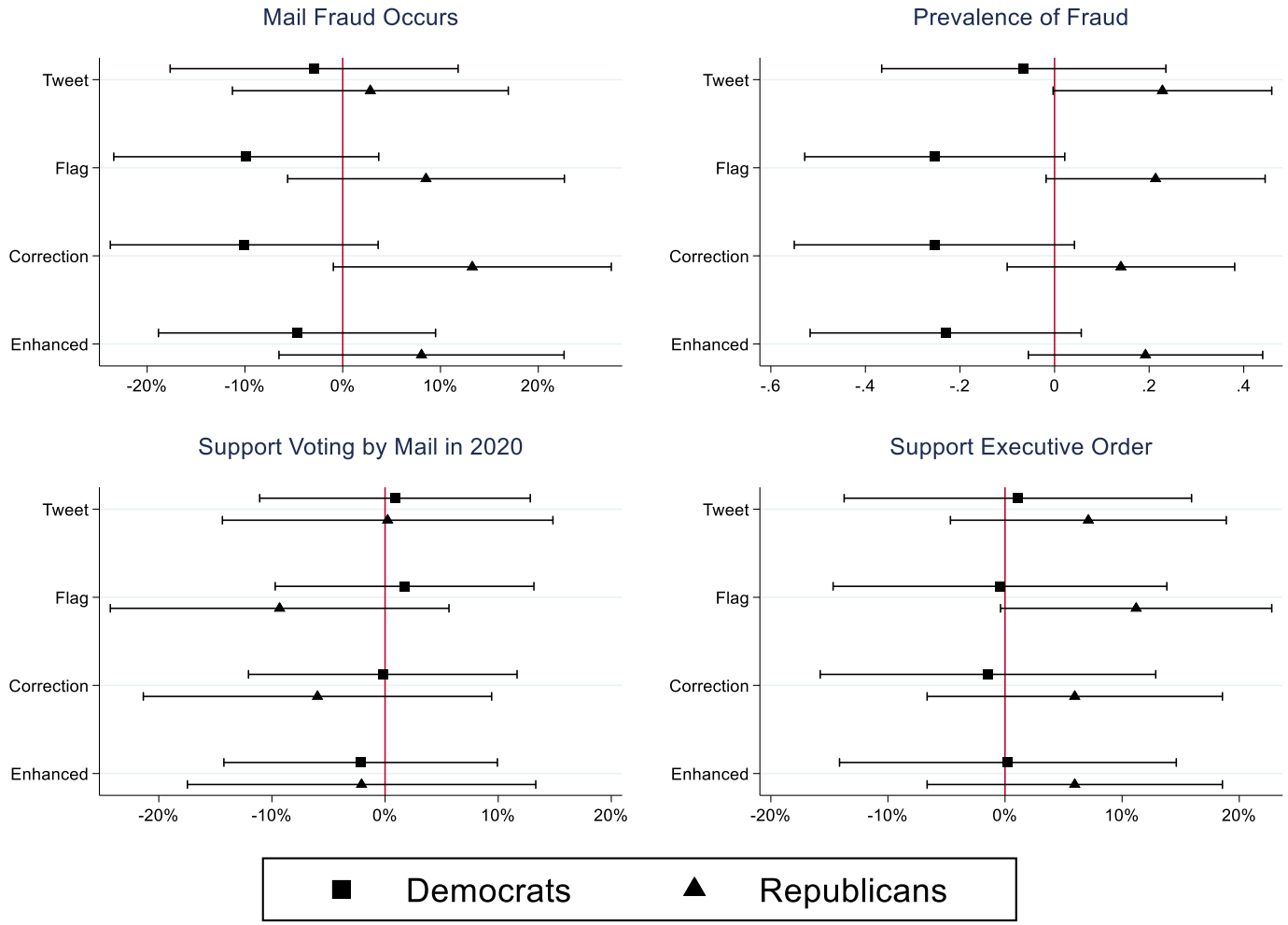


Figure 2: Effect of Tweet and Correction Treatments



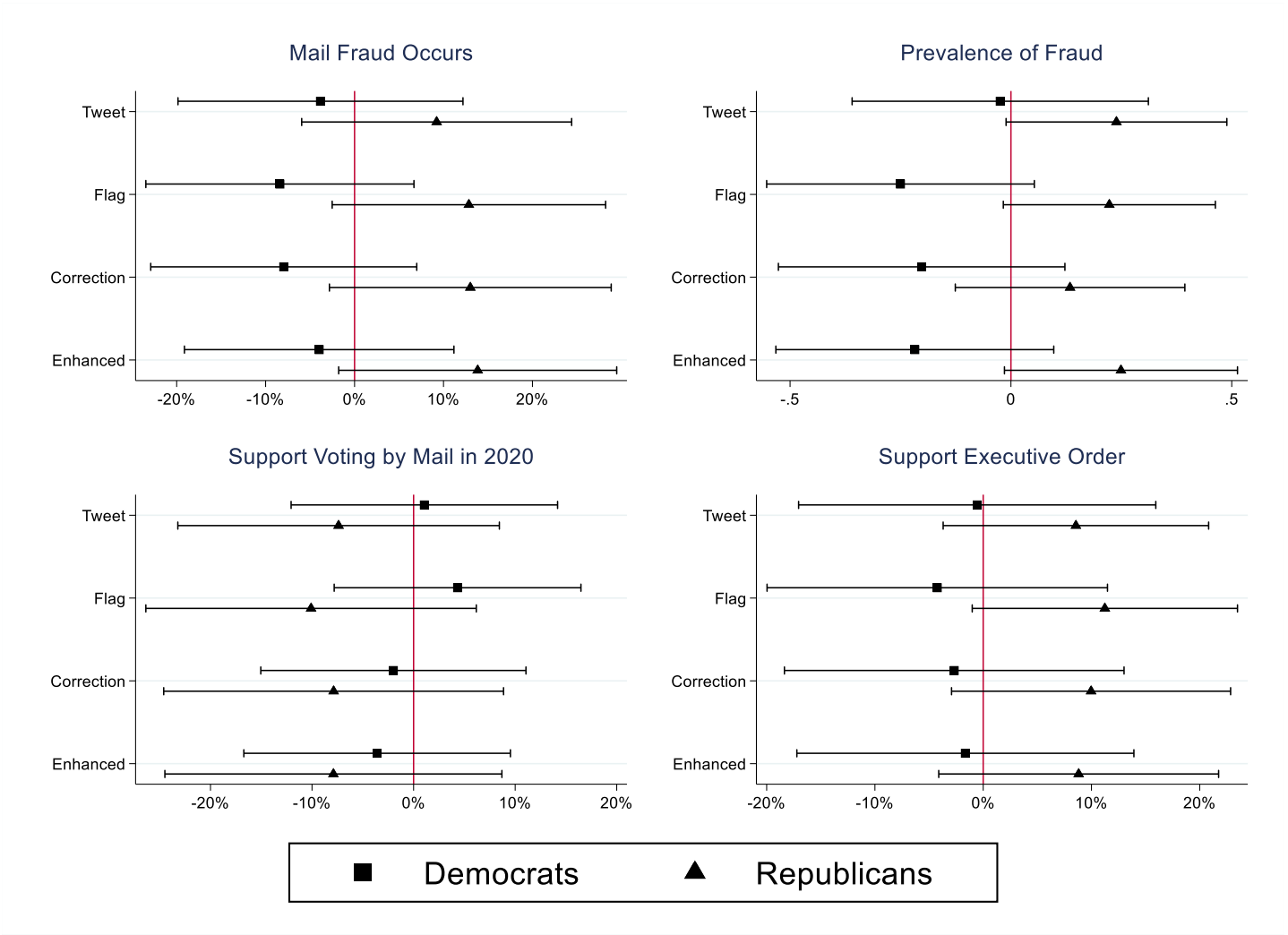
Note: I-bars present 95% confidence intervals about each difference in means (between treatment and control group).

Figure 3: Effect of Tweet and Correction Treatments by Party



Note: I-bars present 95% confidence intervals about each difference in means (between treatment and control group).

SI Figure 1: Treatment Effects by Partisan Groups (Excluding Leaners)



Note: I-bars present 95% confidence intervals about each difference in means (between treatment and control group).

SI Table 1: Comparative Sample Demographics

	Lucid sample	2016 ANES	2018 GSS	US Census
Demographics				
Black	13%	9%	16%	13%
Latino	9%	11%	6%	18%
Female	50%	52%	55%	51%
% College degree	44%	39%	33%	32%
Median age	43 years	49 years	48 years	38 years
Political Characteristics				
Republican	35%	29%	23%	
Democrat	35%	34%	32%	
Ideology (% moderates)	32%	21%	38%	

Note: Partisan figures do not include those who lean toward one party or the other.

SI Table 2: Randomization Checks

	Control	Tweet	Flag	Correction	Enhanced	F-statistic	P-value
Democrat	.36	.37	.42	.43	.42	.83	(.51)
Republican	.44	.46	.41	.40	.38	.82	(.51)
Education	3.90	4.20	3.84	4.05	3.96	1.42	(.22)
Age	44.34	45.48	43.71	45.53	43.70	.57	(.68)
Female	.55	.46	.52	.48	.50	1.06	(.38)
Black	.11	.11	.13	.14	.16	1.06	(.38)
Latino	.10	.09	.10	.09	.10	.08	(.99)
Observations	199	203	208	192	201		

Note: F-tests and p-values are from a one-way ANOVA of the null hypothesis of equal means across the experimental conditions. In no case can we reject the null of equal means, $p < .05$.

SI Table 3: Regression Models Assessing Treatment Effects

	Mail fraud	Electoral fraud	Vote by mail 2020	Support EO
Tweet	-0.04	0.12	-0.21	0.14
	(0.22)	(0.19)	(0.22)	(0.23)
Flag	-0.04	-0.13	-0.43*	0.16
	(0.22)	(0.18)	(0.22)	(0.23)
Correction	0.14	-0.21	-0.23	-0.06
	(0.22)	(0.19)	(0.23)	(0.24)
Enhanced correction	0.09	-0.10	-0.20	0.10
	(0.22)	(0.19)	(0.23)	(0.23)
Democrat	-0.46**	-0.46***	1.36***	-0.63***
	(0.20)	(0.17)	(0.20)	(0.19)
Republican	1.41***	1.07***	-0.33*	1.82***
	(0.19)	(0.17)	(0.19)	(0.20)
Female	-0.24*	-0.16	-0.14	-0.23
	(0.14)	(0.12)	(0.14)	(0.15)
Age	-0.01**	-0.02***	-0.00	-0.01*
	(0.00)	(0.00)	(0.00)	(0.00)
Education	-0.01	-0.07*	0.11***	-0.04
	(0.04)	(0.04)	(0.04)	(0.04)
Black	-0.15	0.30	-0.35	0.29
	(0.22)	(0.19)	(0.23)	(0.22)
Latino	-0.31	0.06	0.20	-0.30
	(0.25)	(0.21)	(0.26)	(0.26)
Constant	-0.01		0.20	0.31
	(0.34)		(0.34)	(0.35)
Observations	1,003	1,003	1,003	1,003

Note: Mail fraud; support for voting by mail in 2020; and support for executive order are logistic regressions. Electoral fraud is an ordered logit regression. Robust standard errors in parentheses. All significance tests are two-tailed.

*** p<0.01, ** p<0.05, * p<0.10

SI Table 4: Treatment Effects by Party

	Mail fraud	Electoral fraud	Vote by mail 2020	Support EO
Tweet	-0.21	-0.30	-1.00**	-0.15
	(0.51)	(0.43)	(0.50)	(0.48)
Tweet * Democrat	0.06	0.13	1.01	0.20
	(0.62)	(0.53)	(0.67)	(0.60)
Tweet * Republican	0.31	0.86*	1.00*	0.56
	(0.59)	(0.52)	(0.58)	(0.61)
Flag	-0.09	-0.42	-1.30***	-0.25
	(0.49)	(0.43)	(0.49)	(0.47)
Flag * Democrat	-0.45	-0.24	1.44**	0.20
	(0.61)	(0.52)	(0.66)	(0.59)
Flag * Republican	0.46	0.91*	0.94	1.00
	(0.59)	(0.51)	(0.58)	(0.62)
Correction	0.33	-0.56	-0.37	-0.69
	(0.49)	(0.43)	(0.51)	(0.49)
Correction * Democrat	-0.86	-0.06	0.36	0.60
	(0.62)	(0.52)	(0.67)	(0.61)
Correction * Republican	0.29	0.86*	0.12	1.05*
	(0.60)	(0.52)	(0.60)	(0.63)
Enhanced correction	0.07	-0.19	-0.52	-0.03
	(0.47)	(0.42)	(0.48)	(0.45)
Enhanced * Democrat	-0.33	-0.49	0.40	-0.03
	(0.59)	(0.51)	(0.64)	(0.58)
Enhanced * Republican	0.28	0.69	0.43	0.37
	(0.58)	(0.51)	(0.57)	(0.60)
Democrat	-0.14	-0.30	0.71	-0.80*
	(0.42)	(0.36)	(0.47)	(0.41)
Republican	1.16***	0.44	-0.83**	1.27***
	(0.40)	(0.35)	(0.41)	(0.41)
Female	-0.24*	-0.16	-0.13	-0.24
	(0.14)	(0.12)	(0.14)	(0.15)
Age	-0.01**	-0.02***	-0.00	-0.01**
	(0.00)	(0.00)	(0.00)	(0.00)
Education	-0.01	-0.06*	0.11***	-0.03
	(0.04)	(0.04)	(0.04)	(0.05)
Black	-0.13	0.32*	-0.36	0.29
	(0.23)	(0.19)	(0.23)	(0.22)
Latino	-0.29	0.07	0.20	-0.31
	(0.25)	(0.21)	(0.26)	(0.26)
Constant	-0.01		0.62	0.59
	(0.43)		(0.44)	(0.42)
Observations	1,003	1,003	1,003	1,003

Note: Mail fraud; support for voting by mail in 2020; and support for executive order are logistic regressions. Electoral fraud is an ordered logit regression. Wald tests show that in the mail fraud model the effects of the flag treatment ($p < .10$, two-tailed test) and correction treatment ($p < .05$, two-tailed test) on Democrats and Republicans are significantly different from one another.

Wald tests also show that the effects of the tweet ($p < .10$; two-tailed test), flag ($p < .01$, two-tailed test), correction ($p < .05$, two-tailed test) and enhanced correction ($p < .01$, two-tailed test) on Democrats and Republicans are significantly different from one another. Robust standard errors in parentheses. All significance tests are two-tailed.

*** $p<0.01$, ** $p<0.05$, * $p<0.10$