Access to social media and support for elected autocrats: Field experimental and observational evidence from Uganda*

Jeremy Bowles[†] John Marshall[‡] Pia Raffler[§]

September 2024

Social media can engage and inform citizens, reducing support for elected autocrats who control traditional media. But this threat encourages such rulers to limit access to social media. In Uganda's electoral autocracy, we estimate effects of both facilitating and limiting social media access on individuals' support for the long-ruling NRM party. Our partial equilibrium experimental results show that three months of subsidized access to social media decreased NRM support among initial supporters. In contrast, VPN users able to maintain access during an election-time social media ban became relatively more supportive of the NRM than individuals less able to circumvent the ban. The mass restriction's distinct consequences reflect broader equilibrium effects: VPN users received unusually pro-NRM content during the ban—partly because it disproportionately reduced critical content production—and the policy elicited backlash among non-VPN users. Our findings illuminate the trade-offs elected autocrats face when deciding whether to restrict access to social media.

*We thank Kate Casey, Gemma Dipoppa, Asim Khwaja, Zoe Marks, George Ofosu, Dan Posner, Amanda Robinson, Xu Xu, David Yang, and audiences at APSA, Boston University, CAPERS, CESifo, EGAP, Harvard, the LSE-NYU Political Economy conference, MPSA, Northwestern, Princeton, Stanford, UCL, University of Wisconsin-Madison, and WGAPE for helpful comments. We are grateful to the J-PAL Governance Initiative for funding the wave 3 data collection and treatment delivery, and to Columbia University's Institute for Economic and Social Research and Policy and Harvard University's Dean's Competitive Fund for Promising Scholarship for funding wave 1 and 2 data collection. We thank Jhonatan Ewunetie and Zafir Vuiya for excellent research assistance, Samuel Olweny and Matrice360 for survey implementation, and the Agency for Transformation for their support. This study received IRB approval from Columbia University (IRB-AAAT4728), Harvard University (IRB20-1935), Mildmay Uganda Research and Ethics Committee (MUREC) (0511-2020), and the Uganda National Council for Science and Technology (UNCST) (SS682ES).

[†]Assistant Professor, Department of Political Science, University College London.

[‡]Associate Professor, Department of Political Science, Columbia University.

[§]Assistant Professor, Department of Government, Harvard University.

1 Introduction

The rise of social media could reshape citizen support for governments across the globe. Compared with traditional media, social media platforms reduce barriers to content production and amplify engagement with online content through horizontal interactions between users (Aridor et al. forthcoming; Zhuravskaya, Petrova and Enikolopov 2020). The implications for citizens' views of their government are ambiguous. On the one hand, social media enables the creation and dissemination of engaging and diverse content that quickly reaches users, which could expose citizens to new political information or perspectives (e.g. Diamond 2010). On the other, large volumes of content, easily-produced misinformation, and individual and algorithmic selection of what gets viewed may instead combine to sort users into political echo chambers that reinforce their prior beliefs (e.g. Sunstein 2018). Evidence from established democracies largely supports the latter perspective, suggesting that social media exposes users to congenial content with limited effects on political preferences (Allcott et al. 2020, 2024; Bail et al. 2018; Guess et al. 2023; Nyhan et al. 2023; Ventura et al. 2023; cf. Fujiwara, Müller and Schwarz 2024; Levy 2021).

But different dynamics may apply in the world's many electoral autocracies—regimes that hold multiparty elections, but where the playing field is heavily skewed in favor of the ruling party (Levitsky and Way 2010). We argue that two distinctive features of these regimes mean that social media content production and consumption patterns are more likely to expose users to novel news and opposition messaging than in established democracies: (i) incumbents exert greater influence over, and censorship of, traditional media outlets; and (ii) online social networks are less politically segregated in these generally less partisan contexts. By providing engaging platforms for suppressed voices that are hard to monitor and silence, social media content is more critical of governments than traditional media and reaches citizens of varying political dispositions. Increasing access to social media should then, all else equal, reduce support for ruling parties in electoral autocracies, especially among supporters of the incumbent likely to have less prior exposure to critical perspectives and hence the most scope to update unfavorably.

Yet, elected autocrats need not take this threat to their popularity as given. Early optimism that social media would serve as a "liberation technology" (Diamond 2010) has since been tempered by the efforts of repressive regimes to control digital media (e.g. Morozov 2012; Tucker et al. 2017). While high-capacity autocrats can precisely censor particular content or accounts (e.g. King, Pan and Roberts 2014),[1] incumbents in the world's many lower-capacity electoral autocracies rely on blunter and less sophisticated censorship tools, such as periodic internet blackouts, imposing

[1] Such governments can also use social media—especially when combined with state-orchestrated censorship—to distract, misinform, and polarize citizens in ways that benefit autocrats (King, Pan and Roberts 2017; Nyabola 2018; Roberts 2018).

platform bans, and raising the costs of using social media. Such partial restrictions are foundational to the survival strategies of modern "informational autocrats," who control information to retain popular support without resorting to repression (Guriev and Treisman 2019; Heo and Zerbini 2024). Nevertheless, autocrats vary significantly in the extent to which they choose to restrict access to social media.

We contend that policies that partially restrict if, and how, citizens use social media *en masse* affect political support in subtle ways beyond the direct effects of limiting their access to online content. First, when they are differentially enforced against opponents—as is typical in electoral autocracies—such policies could achieve a chilling effect by reducing the pro-opposition slant or composition of content producers on social media. Consequently, social media restrictions can reduce critical content by remolding the media market equilibrium. Second, citizens who lose access to social media may sanction the government for making tools they value for news, social networks, entertainment, or business purposes inaccessible. By altering information consumption *and* production decisions, these countervailing content favorability and backlash channels imply trade-offs for elected autocrats considering whether social media will increase their support.

This article examines how facilitating and limiting social media access affects support for Uganda's long-standing incumbent party around its 2021 elections. A canonical electoral autocracy, the National Resistance Movement (NRM) led by President Museveni has ruled since 1986 and exerts substantial control over traditional media sources. By contrast, social media is widely used by opposition-leaning figures—most prominently the main challenger party, the National Unity Platform (NUP). In response to this threat to the NRM's power, the government has limited access to social media by introducing the "over-the-top" (OTT) tax on daily social media use in 2018, levying indirect taxes on mobile data bundles, and—most overtly—imposing a complete internet blackout immediately around the 2021 presidential election and a nominal month-long ban on social media that extended beyond the election.

We leverage complementary research designs to illuminate the consequences for NRM support of increased access to social media in isolation *and* of policies imposing mass restrictions on access. Our field experiment evaluates the former "partial equilibrium" effects by subsidizing access to social media, without meaningfully altering online content or networks. To capture the latter "general equilibrium" consequences of a nationwide policy limiting access to social media, and discouraging the production of regime-critical content, we leverage a difference-in-difference design around Uganda's social media ban. Both designs draw from an original three-wave panel survey of occasional social media users in electorally competitive districts. We augment these analyses with data from Uganda's two most popular social media platforms: seven million Facebook posts from public accounts and groups across the country, which we use to characterize social media content; and high-frequency tracking of panel participants' WhatsApp usage, which we use as a behavioral measure of their social media usage.

In contrast with the echo chambers observed in established democracies, our field experiment finds that increasing individuals' social media access in isolation decreased incumbent support among prior supporters in Uganda's electoral autocracy. Treated respondents received mobile data and had their social media tax paid to facilitate social media use for three months following the electoral campaign; the control group instead received a mobile money transfer, which was both nominally smaller and could be used for any purpose. The intervention increased the likelihood of using WhatsApp on a given day by five percentage points during a period of opposition-leaning content on social media, but only modestly reduced NRM support for our *average* respondent. However, NRM support declined by 0.25 standard deviations among its prior supporters, who became significantly less likely to believe the NRM cares about people like them, felt less warmly toward the NRM, less open to voting for the NRM in the future, and more supportive of opposition parties. This reduction is most pronounced among NRM supporters getting exposed to new critical perspectives and who came to view government performance less favorably. In contrast, non-NRM respondents' views hardly changed. This moderating effect suggests social media could dissipate core support for elected autocrats but also motivate government efforts to limit access.

The findings from the election-time social media ban illustrate the subtler ways that policy restrictions shape social media's consequences for incumbent support. Our difference-in-differences design compares changes in NRM support across respondents who did and did not previously use virtual private networks (VPNs), which enabled individuals to circumvent the ban. Our behavioral data shows that prior VPN users also became five percentage points more likely to use WhatsApp on a given day during the ban than non-VPN users, but—unlike the field experiment—came to view the NRM *relatively positively*. These differences between VPN and non-VPN users are robust to potential violations of the identifying parallel trends assumptions and alternative operationalizations of VPN usage.

The distinct effect of this large-scale government policy appears to be driven by two complementary mechanisms. First, our corpus of Facebook posts shows that—while maintaining an anti-government tilt overall—social media content became relatively less critical of the regime during the ban. This reduction in critical online content available to VPN users, compared to before the ban, appears to reflect both unexpectedly favorable news about election fairness at this time as well as disproportionately less content being produced by regime critics. Suggesting that social media content persuaded users rather than calcified existing views, relatively greater support for the NRM among VPN users was concentrated among respondents whose prior beliefs about NRM governance and Ugandan democracy were least favorable. Second, respondents who lost access to social media due to the ban became relatively less supportive of the regime. This is primarily driven by individuals who lost access to reliable news, rather than those whose livelihoods were harmed.

Together, the content favorability effect on social media users and the backlash effect among citizens who lost access appear to outweigh, in at least in this instance, the effect of reduced exposure to critical content established by the field experiment.

This paper makes two main contributions. First, we provide some of the first experimental evidence from a non-democracy on how "partial equilibrium" access to social media can threaten support for elected autocrats. Unlike the minimal political effects produced by Facebook and WhatsApp in democracies (Allcott et al. 2020, 2024; Guess et al. 2023; Nyhan et al. 2023; Ventura et al. 2023), our evidence is consistent with social media instead exposing individuals to novel regime-critical content in a setting where traditional media is largely under government control. This finding aligns with observational studies showing that regional *internet* expansion more generally has reduced government approval in the Global South (Donati 2023; Miner 2015), particularly in regimes with uncensored internet but high press censorship (Guriev, Melnikov and Zhuravskaya 2021). By establishing these reductions are driven by regime supporters, our results align with prior evidence on the effectiveness of other media forms in new democracies to affect public opinion by exposing voters to counter-attitudinal perspectives (Brierley, Kramon and Ofosu 2020; Conroy-Krutz and Moehler 2015; Lawson and McCann 2005; Platas and Raffler 2021). These results, which imply a moderating effect of social media, contrast with pessimistic findings from advanced democracies on subjective wellbeing (Allcott et al. 2020; Braghieri, Levy and Makarin 2022; Mosquera et al. 2020), social cohesion (Enikolopov et al. 2024), and hate crime and xenophobia (Bursztyn et al. 2024; Müller and Schwarz 2023), and suggest that there may be fewer adverse consequences of increasing access to social media in low-income countries, where citizens' experience with online activity is more limited and of a different nature. As such, our results fit with studies finding that social media can coordinate citizen protests against autocratic governments (Enikolopov, Makarin and Petrova 2020; Steinert-Threlkeld 2017).

Second, our "general equilibrium" findings from a government ban broaden the political costs and benefits to elected autocrats of media censorship. While our experimental results highlight incentives to restrict access to critical content, blanket or partial media restrictions are not always pursued. Prior studies explain that governments may also need the media to mobilize citizens (Gehlbach and Sonin 2014), surveil government officials (Egorov, Guriev and Sonin 2009; Lorentzen 2014; Qin, Strömberg and Wu 2017), or gauge public opinion (Huang, Boranbay-Akan and Huang 2019; Morozov 2012; Qin, Strömberg and Wu 2017). We advance understanding of the censorship trade-off pertaining to social media by showing how partial bans can increase incumbent support by altering the composition of online content production, in line with prior work showing that censorship alters newspaper and TV market equilibria (Kronick and Marshall 2024; Qin, Strömberg and Wu 2018). Conversely, we show that governments also lose support among citizens who lost access to news on social media. This finding complements studies showing that citizens are more supportive of governments that permit access to entertaining and informative television content (Kern and Hainmueller 2009; Kronick and Marshall 2024), and valued consumption goods more generally (e.g. Manacorda, Miguel and Vigorito 2011). By using the unexpected introduction of Uganda's social media ban to illuminate these subtler considerations, our observational study contributes to a small literature leveraging rare opportunities to study institutional change in real-time (Callen, Weigel and Yuchtman forthcoming).

2 Social media in electoral autocracies

How might social media affect support for incumbents in electoral autocracies? We begin by extending competing theoretical logics from established democracies, holding constant the content available online. Then, recognizing elected autocrats' ability to shape access to social media, we move beyond this partial equilibrium analysis to consider the consequences of policies to restrict access on content production and citizen satisfaction—and thus some of the trade-offs shaping the ruling party's decision to censor in the first place.

2.1 Social media and regime-critical information

Social media is distinct from traditional media in two salient ways: through lower barriers to entry for content producers and enabling horizontal interactions between consumers (Aridor et al. forthcoming; Zhuravskaya, Petrova and Enikolopov 2020). Regarding the former, while broadcasting on radio or television entails producing content that appeals to editorial gatekeepers, individuals and organizations can almost costlessly disseminate content on social media platforms and use their algorithms to promote popular content. These low barriers to entry permit a larger and more rapidly-evolving cast of content producers relative to traditional media. Regarding the latter, social media permits engagement with content between citizens. Most obviously, the ability to share, comment on, and approve and disapprove content on platforms like Facebook and WhatsApp enables users to spread content and convey their sentiments to others. This potentially exposes citizens to new information, arguments, and common understandings. Either through direct communication (Enríquez et al. 2024) or by generating common knowledge (Cornand and Heinemann 2008; Morris and Shin 2002), this horizontal engagement can then shape offline social interactions while coordinating users' political views.

How the distinct features of social media platforms interact to shape support for incumbents is theoretically ambiguous. We argue that it depends on whether social media predominantly sorts users into *political echo chambers* or exposes them to *alternative news and perspectives*. In electoral autocracies, we propose that the latter channel is likely to dominate the former, implying reduced support for incumbents.

Evidence from established democracies highlights the possibility of political echo chambers. Because news feeds are individualized, social media could more effectively sort users into political echo chambers than traditional media (Peterson and Kagalwala 2021; Sunstein 2018) or enable users to avoid political content entirely (Prior 2007). Rather than exposing users to new information and diverse opinions, social media platforms would then—either consciously or algorithmically—help users select into groups of individuals sharing like-minded content, a key dimension of which is political in partisan contexts. Such echo chambers may then fail to challenge users' existing political views. Evidence from developed democracies finds that users are indeed largely exposed to congenial online content, with limited effects on political attitudes or behaviors (Allcott et al. 2020; Nyhan et al. 2023). Moreover, social media platforms coordinate offline behaviors around pre-existing views (Bursztyn et al. 2024; Enikolopov, Makarin and Petrova 2020; Fujiwara, Müller and Schwarz 2024; Müller and Schwarz 2023), further solidifying users' existing political opinions.

Extending this logic to electoral autocracies would imply that social media does little to alter political support, or may even entrench support for incumbents. However, its distinct role as an alternative news source is more likely to reduce support for incumbents in these settings for two reasons.

First, ruling parties exert significant control over traditional media sources (Levitsky and Way 2010). Content then tends to favor incumbents because traditional media are owned by the ruling party (Enikolopov, Petrova and Zhuravskaya 2011; Szeidl and Szucs 2021), dependent on state regulation and advertising revenues (Di Tella and Franceschelli 2011), or sufficiently few to be bought off (Besley and Prat 2006; McMillan and Zoido 2004). By contrast, elected autocrats face greater costs of identifying and suppressing regime-critical content producers online. Whereas rulers can relatively easily target specific TV stations or newspapers, and government-influenced media often report selectively to increase support for the ruling party (e.g. Gehlbach and Sonin 2014; Guriev and Treisman 2020), few governments have the capacity to continuously monitor and remove specific social media content or accounts—whether through direct oversight or forcing social media platforms to act as their agents.[2] The critical voices typically suppressed by traditional media then generate an imbalance in partisan slant, with social media content relatively more regime-critical than traditional media.

Second, the impact of social media as an alternative source of regime-critical content is amplified by the relative weakness of partisan attachment in many electoral autocracies. In these settings, while opposition partisans often possess distinct ideological convictions (Carlson 2016; Weghorst 2022), nominal regime supporters—not to mention the large share of non-partisan citizens (Letsa

[2]China's vast and authoritarian state apparatus, enabled by its control over domestic social media platforms, is an exception that censors the production of certain online content (King, Pan and Roberts 2013, 2014).

and Morse 2023)—are often only weakly attached to their partisan identities, the extent of which increases with economic motives (Rosenfeld 2020), a lack of information about the regime (Reuter and Szakonyi 2021), and the homogeneity of their offline social networks (Letsa forthcoming). As highlighted by evidence from new and struggling democracies (Brierley, Kramon and Ofosu 2020; Conroy-Krutz and Moehler 2015; Lawson and McCann 2005; Majumdar forthcoming; Platas and Raffler 2021), the general weakness of partisan attachment implies that many citizens are likely to be more receptive to counterattitudinal information. Moreover, by limiting users' intentional or inadvertent sorting into online social networks by partisanship, it hinders the formation of political echo chambers and their calcifying consequences for political attitudes.

Putting this together, we expect social media to predominantly reduce support for incumbents in electoral autocracies. If critical content is suppressed less on social media than on traditional media *and* networks on social media are not too segregated by partisanship, social media can expose citizens to less favorable information or commentary about government performance, opposition policy proposals and ideologies, or popular discontent. To the extent it is persuasive—because information is novel and credible (see Little 2023) or arguments are compelling—and less favorable toward the government than the information sources people would otherwise rely on, social media will cause citizens to update negatively about the government's competence or alignment with citizen's interests. As we illustrate in a simple model in Appendix A.1, any such negative updating is likely to be concentrated among ruling party supporters, since they hold more favorable prior beliefs about the incumbent party (e.g. Arias et al. 2022; Banerjee et al. 2011; Bhandari, Larreguy and Marshall 2023; Platas and Raffler 2021) or are most attached to the incumbent party for other reasons (Heo and Zerbini 2024). Social media's potential to systematically harm elected autocrats may be further amplified by explicit communication (e.g. Barbera and Jackson 2020; Shadmehr and Bernhardt 2017) and common knowledge among citizens (e.g. Chwe 2000; Kuran 1989).

2.2 Incentives to introduce policies to limit access to social media

The preceding "partial equilibrium" analysis took social media policies and relatively more critical content on social media in electoral autocracies as given. But, if social media primarily serves as a source of content critical of the regime, incumbents are not powerless to resist this challenge to their support. Examples of modern "spin dictators" reshaping traditional media markets to retain support without resorting to targeted repression abound, from Russia to Singapore to Venezuela (Guriev and Treisman 2019, 2022).

Though comprehensive monitoring and targeted suppression of regime-critical content on social media is often impossible, particularly in low state capacity settings, incumbents often employ blunter tools targeting consumers' access to such platforms and periodically sanction producers. In extreme cases, this entails banning social media platforms entirely (Miller 2022). Subtler partial censorship strategies increase the costs of accessing social media by imposing taxes (on social media use or mobile data), slowing websites or platforms down, or limiting access to censored platforms to citizens with VPNs (e.g. Boxell and Steinert-Threlkeld 2022; Roberts 2018). Because such restrictions are generally enforced by a small number of locally-based telecommunications companies, over which governments can easily exert influence, they are more feasible to administer.

We argue that elected autocrats can sustain their support by leveraging these subtler restrictions on access to social media. While incumbents may have complementary motivations for restricting access, we focus on substantiating some of these trade-offs in terms of shaping support for the regime—which may often align with considerations like constraining collective action or the capacity to document election fraud.[3] In particular, we highlight how incentives to censor social media access are likely to be influenced by the "general equilibrium" impact of restrictions on content consumption and production.

Compared with unrestricted access to social media, partial bans could increase support for incumbents through several informational mechanisms. First, social media restrictions *limit exposure* to critical online content and coordinated responses to it. If truthful revelations or persuasive opposition commentary are more prevalent on social media than alternative information sources, like traditional media or in-person discussions, reducing access to social media could maintain support for the ruling party (Gehlbach and Sonin 2014; Guriev and Treisman 2020; Shadmehr and Bernhardt 2015). Partial restrictions that segment media markets may even be optimal for incumbents if they prevent persuadable regime supporters from encountering truthful content but still allow regime opponents to be persuaded by it and become more favorable toward the government when good news is reported (Heo and Zerbini 2024).

Second, partial restrictions to online freedoms may alter what social media content is produced. In particular, online content might become more favorable toward incumbents if restrictions target critical content producers (or are only expected to be enforced against opponents) or reduce demand for their content among their reconstituted audience. Furthermore, reducing competition for online audiences may in turn allow pro-government accounts to become more favorable without losing much of their target audience (Kronick and Marshall 2024). These media market equilibrium effects highlight how censorship could expose users to *more favorable content* without resorting to blanket bans, and may thus affect support for incumbents beyond limiting access to critical content.

However, imposing social media restrictions could also lose political support or mobilize op-

[3] Although independent media could also serve to inform central governments about public opinion or bureaucrat performance (Egorov, Guriev and Sonin 2009; Huang, Boranbay-Akan and Huang 2019; Lorentzen 2014; Qin, Strömberg and Wu 2017), there are often strong incentives—especially ahead of elections—for incumbents with the capacity to limit media freedoms to censor critical media sources.

position. First, social media may be a valued source of objective news, social interaction, entertainment, consumer purchases, and even business opportunities. Limiting access may thus result in *backlash*. Indeed, Hugo Chávez's decision not to renew Venezuela's leading TV channel's public broadcast license cost him votes among viewers who lost access to popular entertainment content and informative news (Kronick and Marshall 2024). Conversely, access to popular West German television increased support for the Soviet Union in East Germany (Kern and Hainmueller 2009). Second, regardless of whether they themselves lose access to social media, citizens who value democratic freedoms or draw inferences about the government's type may disapprove of censoring governments (Gläßel and Paula 2020).[4]

These forces highlight the countervailing incentives facing incumbents concerned about maintaining broad-based support to restrict social media access. While these incentives exist across regime types, the trade-off is especially relevant in electoral autocracies. On the one hand, closed autocracies' more limited sensitivity to backlash renders the incentives to restrict access more straightforward. On the other, the more modest relative partisan imbalance in online slant and higher electoral costs of censorship limit incentives for incumbents in established democracies to restrict access. Figure A1 shows electoral autocracies have similar *capacities* to shut down internet or social media access, or filter access to particular websites, as closed autocracies. But, consistent with their potentially ambiguous trade-off, in practice they restrict access at rates falling between closed autocracies and democratic regimes.

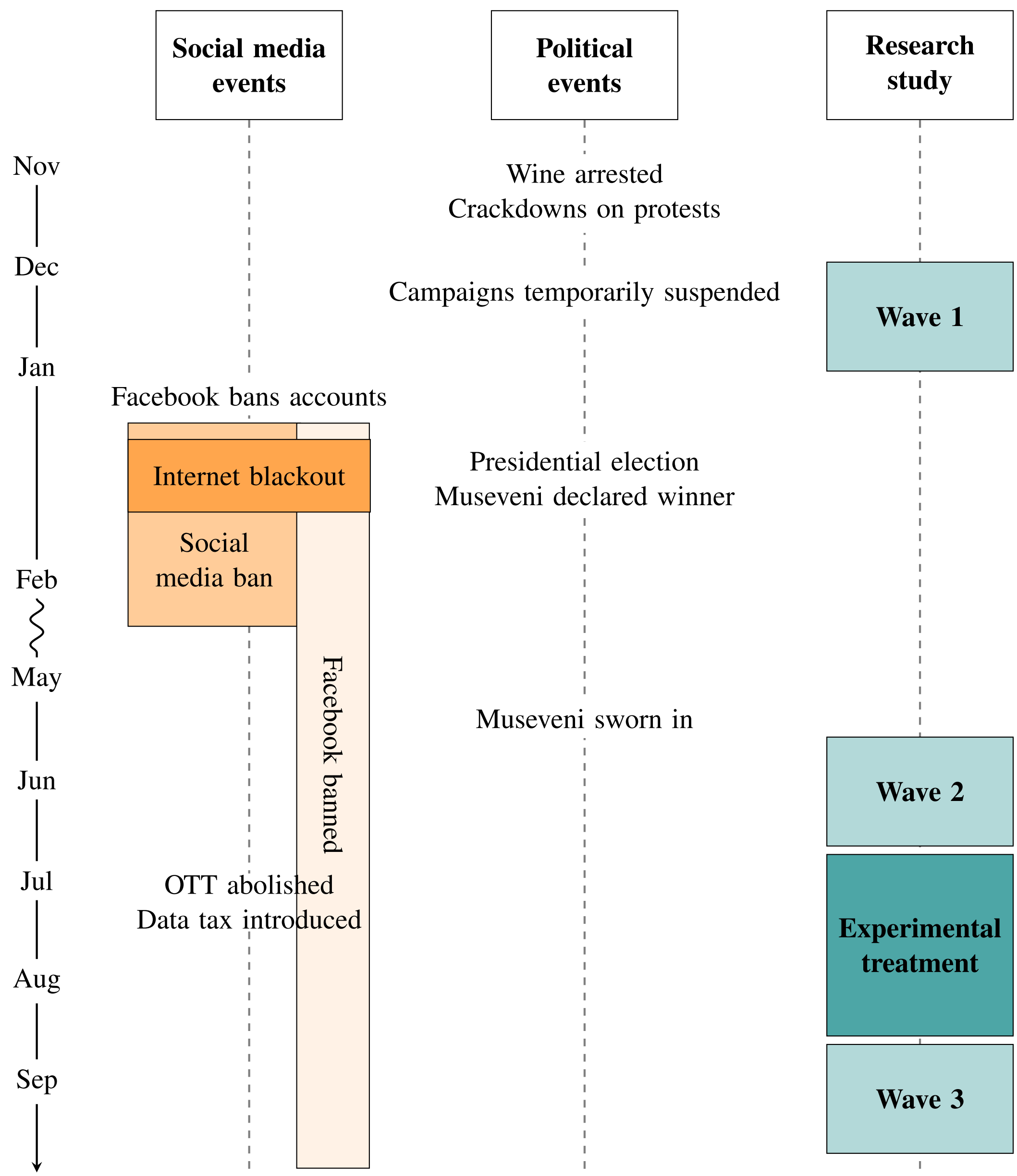
The trade-offs introduced by the potential for mass policies to reshape content production but also upset citizens are captured by our empirical focus on comparisons across citizens who do, or do not, lose access to social media. While all citizens are likely to perceive signals of undemocratic tendencies, our simple model in Appendix A.1 clarifies conditions under which government policies restricting individuals' access to social media may change relative support for ruling parties. Reflecting theoretical ambiguity surrounding which effect dominates in a particular context, we expect incumbent support to increase (decrease) among citizens who lose access relative to those who do not when the positive effects of limiting access to regime-critical content dominates (is dominated by) the broader effects of censorship to induce both more government-friendly social media content and backlash among individuals losing access to social media.

3 Media and politics in contemporary Uganda

This section provides an overview of electoral politics, media consumption, and social media content in Uganda. Figure 1 charts key events during our study period relating to politics, social media

[4] Others might view this as signaling the government's strength or resolve to repress opposition (Simpser 2013).

Figure 1: Timeline of events during research study



access, and data collection.

3.1 Electoral context

Uganda has continuously been ruled by Yoweri Museveni and his National Resistance Movement (NRM) party since 1986. The most recent presidential elections were held on January 14, 2021. Museveni faced his most credible opposition from Robert Kyagulanyi Ssentamu, nicknamed Bobi

Wine, a rapper-turned-MP with broad support among younger voters. Kyagulanyi represented the National Unity Platform (NUP), and rapidly surpassed the Forum for Democratic Change (FDC) as the leading opposition party.

Since the 2021 elections coincided with the COVID-19 pandemic and restrictions on public gatherings, the Electoral Commission dictated that the campaigns would follow a "scientific" model using broadcast and online media to appeal to voters, rather than the typical holding of mass rallies (Uganda Communications Commission 2021). Nevertheless, enforcement of in-person campaigning rules was heavily imbalanced. Whereas NRM rallies remained common, opposition rallies were violently disbanded (Freedom House 2021). In November 2020, Kyagulanyi was arrested for violating campaigning restrictions, sparking widespread protests whose suppression resulted in 54 reported deaths (Amnesty International 2021). In this repressive context, many voters likely anticipated further crackdowns and malpractice at election time.

Museveni ultimately won the presidential election with 58% of the official vote, followed by Kyagulanyi with 35%, and was inaugurated for his sixth term in May 2021. In the parliamentary elections, the NRM won the most seats (64%), followed by independents (14%), the NUP (11%), and the FDC (6%). The United States described the elections as "neither free nor fair."[5]

3.2 Traditional and social media consumption

Traditional media remains Uganda's most common source of news. In 2019, 58% of people reported listening to news on the radio at least a few times a week, followed by 34% for television and 12% for print newspaper (Afrobarometer 2019). The regime exercises considerable control over such media (Freedom House 2021); accordingly, only 23% of citizens believe the news media is completely free to report and comment on the news (Afrobarometer 2019). Prior to the election, journalists were arrested for hosting opposition candidates on their shows, a radio station was raided, journalists were prevented from covering opposition rallies, and foreign journalists were denied accreditation (US Department of State 2020). Likely owing to this sanctioning, our survey respondents perceived radio and TV as notably more supportive than critical of the NRM.

However, social media has become increasingly popular and constitutes the vast majority of internet usage. As in most sub-Saharan African settings, access is almost entirely through cell phones, with 52% of Ugandans having mobile internet connections in late 2020. By 2019, Afrobarometer (2019) shows that 14% of Ugandans used social media to obtain news at least a few times a week; moreover, 88% of users believe that it informs people about current events, and social media users were regarded as less likely to spread false information than government officials

[5]Press statement by Secretary Blinken, April 16, 2021; www.state.gov/imposing-visa-restrictions-on-ugandans-for-undermining-the-democratic-process.

or political parties. Respondents in our pre-election survey of occasional social media users reported spending five times more time on social media applications in a normal week than browsing websites. Facebook and WhatsApp are by far the most popular platforms, with 79% and 78% of our respondents reporting using them respectively; with only 17% reporting using Twitter.[6]

Ugandan social media content is fairly political. Among the Facebook users in our sample, Figure A2 shows that 71% viewed getting political news as a main reason for using social media and 25% cited discussion of current events. Among WhatsApp users, these figures are 53% and 26%, respectively. Importantly, as in many countries in the Global South, WhatsApp is not just a private messenger app, but also a form of mass communication via groups of up to 256 users. Further, networks are quite politically heterogeneous: 84% of our respondents stated that all, or most, of those with whom they discuss politics held views which varied from their own.

In contrast with traditional broadcast media, social media content is generally more critical of the president than supportive. Around the election, this was driven by the NUP's extensive use of online platforms to reach its young and urban support base—and perhaps amplified by its inability to use in-person campaigning and limited access to broadcast media. As we describe systematically in Section 3.4, this led to a strongly pro-opposition slant on social media.

3.3 Access to social media

Perhaps unsurprisingly in light of this slant, the government has employed various tools to limit citizens' access to social media. Efforts to increase the cost of access started in 2018, when the "over-the-top" (OTT) tax was introduced. Officially motivated as raising revenues and reducing exposure to "gossip" online, the OTT required users to pay 200 shillings ($0.055) per day to access social media platforms, including Facebook, WhatsApp, Instagram, and Twitter. While citizens could evade the tax by using VPNs, doing so was slower and more bandwidth-intensive (Pollicy 2020). Due to its limited revenue generation, the government replaced the OTT tax with a 12% tax on mobile data in July 2021, on top of the existing 18% VAT. These taxes contribute to very high data costs, with 1GB of mobile data costing 8% of a Ugandan's average monthly income (A4AI 2019).

Since social media users are younger, more urban, and more opposed to the NRM, these policies are widely seen as tools intended to undermine opposition support (Namasinga and Orgeret 2020) by limiting social media usage (Boxell and Steinert-Threlkeld 2022). Civil society organizations decried the OTT tax, for example, as "a clear attempt to silence dissent, in the guise of raising government revenues" (Amnesty International 2018). The OTT tax is listed by 55% of our sample

[6] Traditional and social media consumption is relatively balanced in this sample, with 80% reporting listening to the radio while 72% report watching TV.

as preventing them from using social media more, while 67% listed the cost of data.

Around the election, the government imposed blunter tools to control access to social media. In early January, Facebook removed a network of hundreds of government-linked accounts for engaging in "coordinated inauthentic behavior" promoting the NRM and denigrating the NUP, with Twitter following suit. On January 12, nominally to avoid the spread of misinformation, the Uganda Communications Commission (UCC) announced a ban of all social media platforms, including WhatsApp, Facebook, and Twitter. Enforced by demanding that internet service providers block access to the platforms regardless of users' OTT tax payment, the ban was lifted on February 10 except for Facebook, which remains officially blocked. However, many individuals—including many government officials—used VPNs to maintain access during the ban. Because it could be circumvented, we characterize the "ban" as introducing significant friction and uncertainty about the consequences of using social media. On the eve of the election, the UCC shut the internet down completely. Internet access resumed five days later, shortly after Museveni was declared the election's winner.

3.4 Social media content over time

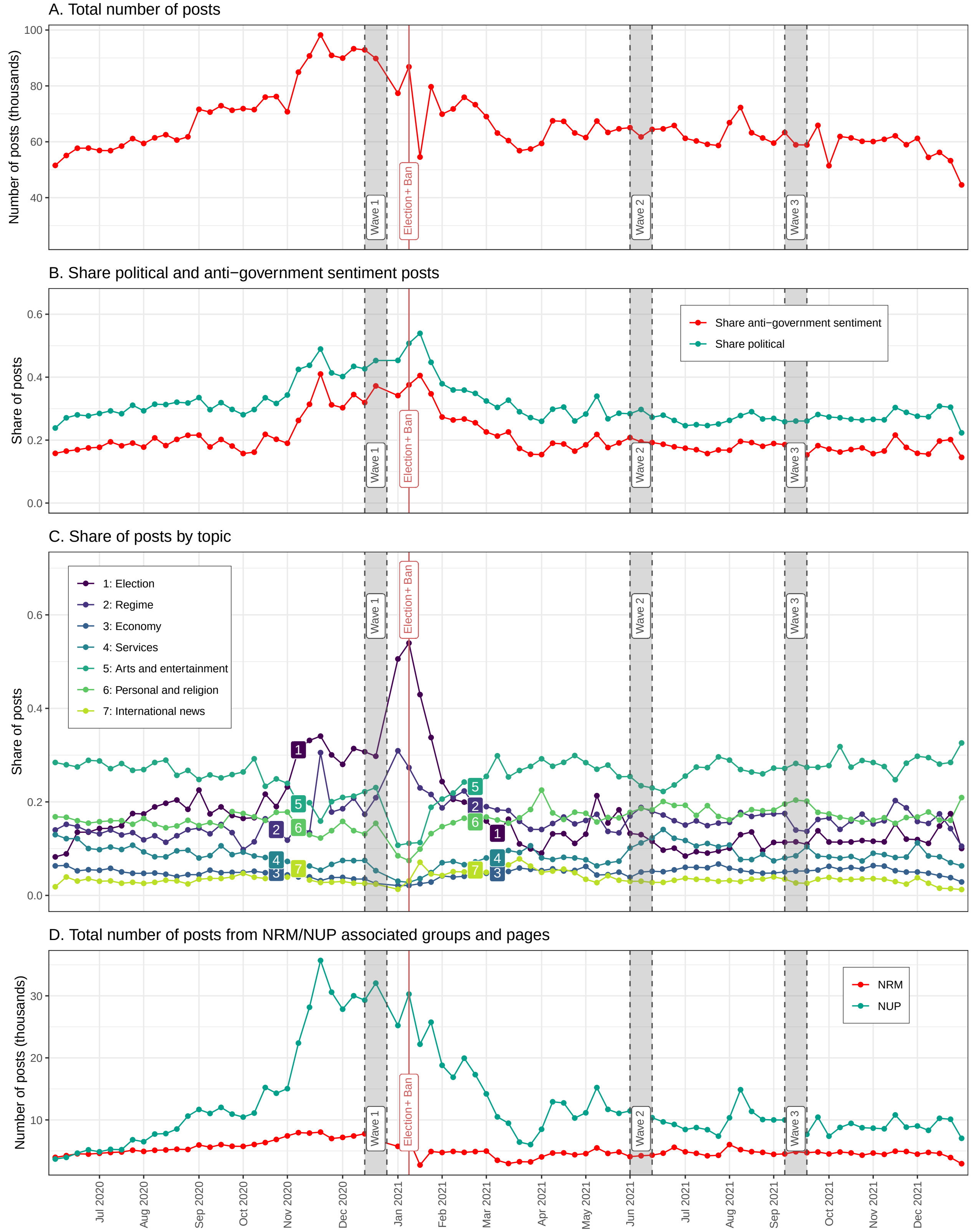
Finally, we describe social media content during the study period. We focus on Facebook, for which many posts from relatively popular accounts are publicly available. Respondents' similar reasons for using WhatsApp—where private content is encrypted—suggest that it played a similar political role in this context (see Figure A2).

To characterize content on Facebook, we collected a corpus of 6.95 million publicly-accessible posts from 12,521 distinct Facebook pages in the Crowdtangle database between June 2020 and December 2021. The corpus comprises the universe of public Facebook pages self-recorded as being in Uganda and a large curated set of groups and pages pertaining to Ugandan politics; Appendix A.2.1 fully details its composition. We then use a combination of GPT and BERT to classify every English-language post (83% of the sample) in terms of (i) whether it is about Ugandan politics, (ii) whether its sentiment is anti-government, neutral, or pro-government, and (iii) the substantive topic of the post. For a subset of the corpus taken from salient political pages and groups, we further label the partisanship of the account. Appendix A.2.2 details our classification process.

Figure 2 aggregates posts by week throughout our study period, showing that the high initial level of political content as well as the strongly anti-government slant and composition of producers fell during the social media ban and after the election.[7] Panel A first documents that the quantity of posts of all kinds peaked in the pre-election period, but dropped sharply after the social media ban

[7] Similar patterns emerge when aggregating the total interactions with posts (see Figure A4) or restricting to the subset of relatively political pages and groups (Figure A5).

Figure 2: Facebook posts during study period



Notes: See Appendix A.2 for information on the corpus and its classification.

was imposed. Consistent with extensive circumvention of the ban, substantial posting continued even once Facebook could only be accessed by VPN.[8]

Turning to the content of posts, Panel B shows that the *share* of posts classified as pertaining to politics also peaked immediately before the election, when around half of all posts were classified as being political. The share of posts classified as containing anti-government sentiment similarly peaked just before the election, reaching a maximum of nearly 40% of all posts—seven times more than the pro-government share of posts. Panel C shows that the share of posts about elections and regime-connected topics relating to government institutions, civil rights, protests, the military, and corruption were common throughout the campaign, spiked before the election, and remained common even after President Museveni's inauguration, while posts about the economy, public services, and especially international news were comparatively rare. During the social media ban, such political content was largely displaced by posts about arts, entertainment, religion, and personal issues. Panel D underscores the dominance of NUP-associated accounts relative to NRM-associated accounts prior to the election, along with their especially precipitous fall in posting shortly afterwards. Figure A6, which instead categorizes the full set of accounts according to the share of their posts classified as anti-government prior to the ban, shows that the most strongly anti-government accounts were those which reduced the political tone of their posts most.

4 Data collection

Our analyses draw from an original three-wave panel survey conducted during and after the 2021 election campaign. We first explain our sampling strategy before introducing our survey and behavioral data sources.

4.1 Sampling

We sought to recruit participants for whom accessing information on social media could be politically salient and who already used social media, but use it sufficiently irregularly that they could be induced to do so more frequently. To reach this population, we selected 11 districts—from all regions of Uganda—where the NRM received 40-60% of votes in the 2016 election. Within each district, shown in Figure 3, we sampled participants from peri-urban trading centers (TCs) on the fringes of urban localities with good 3G internet reception. Our sampling frame of 4,399 potential respondents from 135 TCs was constructed by asking "seeds" in each TC to provide contact details for potentially eligible individuals; Appendix A.3 provides further details.

[8]See Figure A3 for a day-level equivalent around the ban.

Figure 3: Sampled districts and trading centers

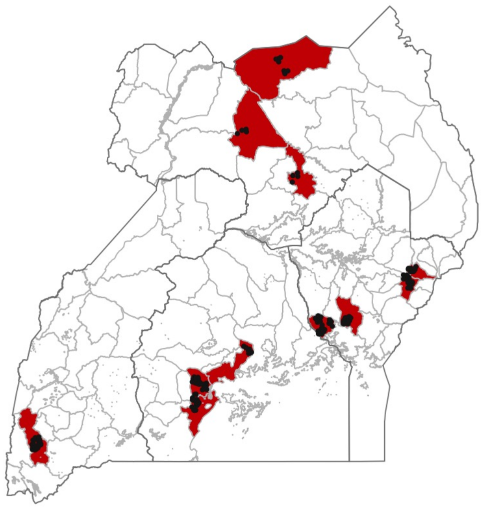
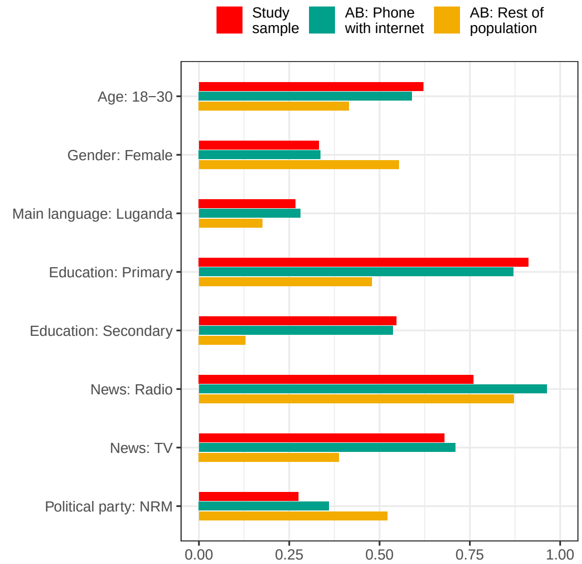


Figure 4: Sample characteristics



Notes: The sampled districts in Figure 3 comprise Mpigi, Kalungu, Masaka in Central region; Iganga, Jinja, Mbale, Sironko in Eastern region; Gulu, Lamwo, Lira in Northern region; and Rukungiri in Western region. Figure 4 compares mean demographic characteristics among our baseline survey sample with adult Ugandans from Afrobarometer (2019).

Our research team called 3,710 potential respondents, of which half met our eligibility criteria: (i) aged 18-50; (ii) possessing a cell phone able to access social media platforms; and (iii) reporting using social media apps three or fewer days in the last week. Ultimately, 1,542 eligible individuals—a potentially consequential class of the electorate likely to become intensive social media users as access and usage increase—completed the baseline survey.

The baseline sample approximates our target population within Uganda. Using nationally representative data from Afrobarometer (2019), Figure 4 shows that our sample matches the average age, gender, language, education, traditional media consumption behaviors, and partisanship of Afrobarometer respondents possessing an internet-accessible phone remarkably well. Those without an internet-accessible phone are older, less educated, more likely to be female, and more favorable to the NRM.

4.2 Panel survey data

We surveyed our panel three times over almost a year. The baseline survey (wave 1) was administered in December 2020 and early January 2021. Wave 2 was enumerated between late May and early June 2021. We successfully re-interviewed 1,310 (85%) wave 1 respondents and added 145 eligible participants we had been unable to reach for wave 1. Wave 3 was administered in September 2021, successfully resurveying 1,389 (95%) of the 1,455 respondents who completed wave 2.

All surveys were conducted via telephone, given COVID-19-related health risks associated with in-person enumeration.[9] Appendix A.5 provides a discussion of the ethics of our data collection and interventions.

Each survey wave measured our core outcomes—support for the NRM and opposition parties—in three ways. First, we asked respondents which party they believed to care most about the welfare of Ugandans; for our outcomes, we consider the incumbent NRM and pool together opposition parties.[10] Second, we used a feeling thermometer to gauge how warmly respondents felt about the NRM and opposition parties on an 11-point scale ranging from 0 (very cold) to 10 (very warm). Third, we asked how open respondents would be, on a five-point scale, to voting for NRM and opposition candidates for a generic political office in the future. Political constraints prevented us from asking directly about presidential vote choice, but openness to voting for a party closely correlates with vote choice in local elections (Platas and Raffler 2021). We aggregate these measures to construct inverse-covariance weighted (ICW) indices (Anderson 2008) capturing support for the NRM and opposition parties.

A key potential moderator of social media's effects is prior partisanship. We measure NRM partisanship using an indicator for respondents who reported both feeling more warmly towards the NRM and being more open to voting for the NRM than any opposition party. This designates 27% of respondents as NRM supporters, with non-NRM supporters split between opposition supporters and non-partisans. Figure 4 shows that this proportion aligns with the share of NRM partisans among individuals with internet-connected phones across Uganda.

We further elicited respondents' demographics, social media consumption (and whether accessed by VPN), and perceptions of government. Summary statistics are provided in results tables and Table A7.

4.3 Social media usage data

For a behavioral measure of social media usage, we collected publicly-accessible data on the extent—but not content—of respondents' WhatsApp usage. As noted above, WhatsApp usage correlates with the self-reported use of Facebook and other platforms in our sample, both in terms of usage rates and respondents' main reasons for usage (see Figure A2).[11] We audited the WhatsApp status of the 60% of baseline respondents whose phone numbers were linked to active accounts

[9]Our initial intervention comprised paying respondents' OTT taxes after the wave 1 survey and shortly prior to the election. However, the social media ban and internet blackout negated this treatment, leading us instead administer a similar intervention following the wave 2 survey.

[10]As some respondents noted, a lack of experience with opposition government rule may have prevented some opposition supporters from selecting opposition parties.

[11]We cannot track Facebook activity because we did not record account URLs for logistical reasons.

between four and five times per day throughout the study.[12] This enables us to construct a panel of respondent-by-date measures of: (i) whether the respondent had been "last seen" using WhatsApp on a given day; and (ii) the number of distinct timestamps for which the respondent had been "last seen" using WhatsApp on that day.[13] Table A7 shows that the subsample for whom we were able to audit WhatsApp statuses throughout the study was modestly younger, better educated, less favorable towards the NRM, and used more social media at baseline compared to those not audited. Among those we audited, 22% used WhatsApp on the average day during our study.

5 Field experiment

To assess whether facilitating social media access depressed citizens' support for the NRM, we conducted a field experiment in mid-2021, following President Museveni's inauguration. Our treatment subsidized randomly-selected individuals to use social media, but not for clusters of users simultaneously. The intervention thus captures a partial equilibrium effect, by altering individuals' access, without meaningfully changing the supply of online content, peers' social media access, or government social media policies.

5.1 Experimental design

After completing the wave 2 survey, we randomly assigned respondents to receive a financial incentive to increase their social media use for three months. Treated participants were compensated for taking the survey with payments alleviating the main financial barriers limiting social media usage: the OTT tax and mobile data costs. These payments varied slightly by phone network, as detailed in Appendix A.4.1, but comprised data bundles of 400-500MB per week plus OTT tax payment in June, followed by weekly data bundles of 400MB or 500MB throughout July and August (after the OTT tax was replaced by a mobile data tax). Respondents were free to use their mobile data as they liked but, given social media constituted the main use of the internet in Uganda and respondents' stated financial barriers to access, we anticipated the incentive would principally increase social media usage. Since we did not control what participants consumed, the social media content they were exposed to during the intervention reflects individual and network preferences as well as platform algorithms. Figure 2 shows that political content on Facebook in general was both common and slanted against the government during the intervention period.

[12] Of linkable accounts, 90% had publicly-viewable WhatsApp statuses—the default within WhatsApp. Our measure likely underestimates actual WhatsApp usage because some respondents have multiple SIM cards.

[13] Because we audit every phone number multiple times a day, measure (i) is an accurate measure of daily WhatsApp usage. The upper bound for measure (ii) is the number of times we audit that number on a given day, and thus only captures limited intensive margin variation.

To limit the risk of differential attrition, individuals assigned to the control condition were instead compensated with a flexible mobile money transfer of UGX 6,000. Control participants could use this smaller transfer towards social media access, with the average respondent reporting using around half their mobile money to buy data. Both experimental conditions were subtly communicated as tokens of appreciation for study participation, meaning participants were unlikely to be aware of their treatment condition. In contrast with the social media ban we examine later, this intervention was not attributed to government policy decisions and affected only a tiny fraction of social media users.

Using wave 1 survey data, treatment conditions were block-randomized prior to wave 2 enumeration (see Appendix A.4.1 for details). Table A8 shows that wave 3 participants assigned to treatment, both overall and by partisanship, are statistically indistinguishable from those assigned to control in terms of demographic characteristics and their pre-treatment attitudes. Table A9 demonstrates there is no differential attrition between wave 2 and wave 3, either overall or by partisanship, with low attrition rates of around 5%.

To evaluate how subsidizing social media access affects political support, we use pre-registered OLS regressions of the following form to estimate average treatment effects (ATEs) in our wave 3 survey data:[14]

$$Y_i^{post} = \tau Treatment_i + \alpha \mathbf{Y}_i^{pre} + \beta_b + \gamma_e + \varepsilon_i, \tag{1}$$

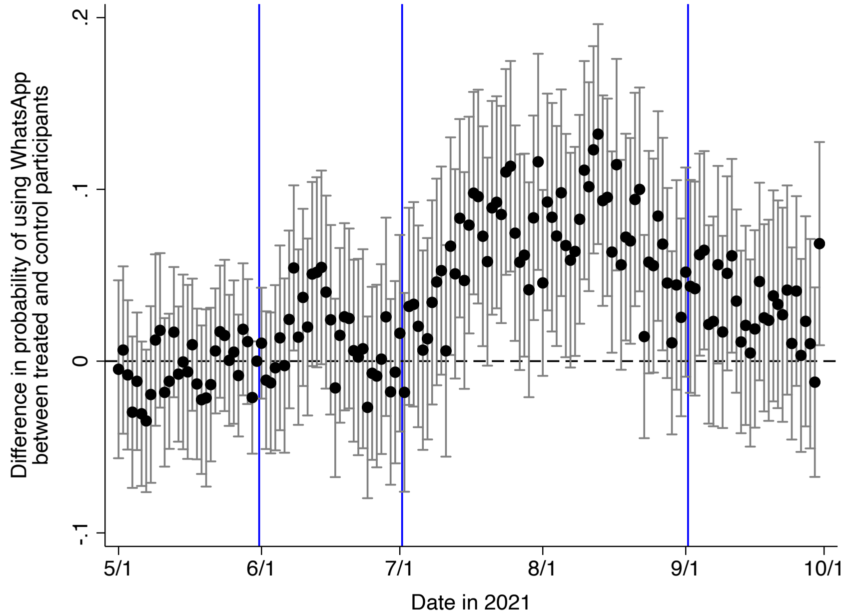
where Y_i^{post} is a wave 3 outcome, $Treatment_i$ indicates receiving our treatment, $\mathbf{Y}_i^{pre}$ is a vector of pre-treatment outcomes (where we use both wave 1 and 2 survey responses, where available), β_b are randomization block fixed effects, and γ_e are wave 3 enumerator fixed effects. Following Belloni, Chernozhukov and Hansen (2014), an auxiliary specification adds a LASSO-selected vector of predetermined covariates for robustness. We use robust standard errors for inference, reflecting the individual-level randomization.

Since non-NRM partisans are likely to have already been exposed to more regime-critical information than NRM partisans, our pre-analysis plan further registered our expectation that an exogenous increase in social media access, relative to other news sources, would have a stronger effect among NRM supporters. To test this, we add our measure of partisanship described above to saturated interactive specifications that estimate conditional average treatment effects and test for differential effects across NRM and non-NRM partisans.[15]

[14]Appendix A.4.3 describes our pre-analysis plan, including explaining several deviations with minimal consequences.

[15]In the specifications with LASSO-selected covariates, we include selected interactive covariates and all their lower-order terms following Blackwell and Olson (2022).

Figure 5: Differences in daily use of WhatsApp, by treatment assignment



Notes: Estimates are from equation (1), where the outcome is daily use of WhatsApp. The baseline category is May 31, 2021. Treatment begins around June 1, 2021. Revised weekly treatment begins around July 1, 2021. Treatment ends around September 1, 2021. All bars indicate 95% confidence intervals.

5.2 Effects on social media usage

We first assess the extent to which the intervention affected social media usage. To capture treatment effect dynamics using our high-frequency behavioral WhatsApp data, we estimate a panel version of equation (1) in Figure 5, where the outcome is whether a given respondent was "last seen" using WhatsApp on a given date.[16]

The results demonstrate that subsidizing social media use significantly increased WhatsApp usage during the treatment period. We distinguish the month of June, in which we sent most of our sample a single large data transfer (without reminders) and hence treatment effects dissipated quite quickly, from July and August when we sent smaller weekly data transfers (with notifications), which generated more sustained treatment effects. The persistence of modest, if diminishing, effects after the conclusion of the treatment in September suggests that social media use activates demand for further use, in line with prior studies (e.g. Chen and Yang 2019).

Pooling across the intervention period, column (1) of panel A of Table 1 shows that treated participants were 5.6 percentage points more likely to use WhatsApp on a given day during the

[16]These regressions include individual-level fixed effects and date fixed effects; standard errors are clustered by respondent. Self-reported social media use data from our surveys is far noisier than our behavioral data (see also Nyhan et al. 2023), likely due to the difficulty of accurately reporting the number of hours spent on social media in a given week up to four months earlier.

treatment period, relative to an average probability among the control group of 0.16 ($p < 0.01$). Column (2) shows that this 35% increase is robust to covariate adjustment. Column (1) of panel B further shows that the number of audit windows within a day that WhatsApp use was registered increase by 0.09 ($p < 0.01$). Columns (3)-(6) report similar effect magnitudes across NRM supporters and non-NRM supporters, respectively. Because our sample comprises a greater share of non-NRM supporters, who we were also able to audit at slightly higher rates, those estimates are more precise.

The audit data also suggest that many treated respondents did not significantly alter their WhatsApp use. To characterize treatment heterogeneity, we use a causal forest to predict individual-level treatment effects on WhatsApp usage (Wager and Athey 2018).[17] The results in Figure A10 suggest that the subsidy particularly increased usage among somewhat less educated participants, those who reported greater prior social media usage, and individuals less supportive of democratic principles. But, in general, treatment effects are relatively similar across subgroups.

These increases in WhatsApp usage likely understate the intervention's effects on overall social media usage among infrequent social media users. First, while WhatsApp usage correlates strongly with usage of other platforms, our treatment may have been more useful in facilitating access to more data-intensive platforms like Facebook that also depend less on friends' connectivity. Second, our measures do not fully capture the *intensity* of participants' usage, since our intensive margin measure only audits usage every five to six hours. Regardless, our moderate-sized "first stage" falls between full deactivation studies (e.g. Allcott et al. 2020, 2024) and nudge-like interventions without monetary incentives (e.g. Levy 2021), but contrasts with prior work by *increasing* access rather than reducing it.

5.3 Partisan-moderated effects of access to social media on NRM support

Having established that treated respondents were more likely to access social media, we next examine changes in support for the ruling NRM party and opposition parties. Table 2 reports our estimates of average and conditional average treatment effects for our ICW indexes of party support and its three constituent items.

Our findings for the full sample of respondents document negative but limited average treatment effects. Column (1) of panel A reports a modest and statistically insignificant negative effect on our index of support for the NRM ($p = 0.23$), with negligible effects on support for opposition parties shown in column (5) ($p = 0.78$). Panel B finds no evidence of effects on beliefs about which party

[17]Our outcome is the difference in the share of days a respondent used WhatsApp during the treatment period relative to before. For predictors of heterogeneity, we considered predetermined outcomes, the individual indicators comprising them, fixed effects by trading center, age, and gender, and respondents' religious, economic, and educational characteristics (defined in either wave 1 or 2).

Table 1: Field experimental treatment effects on daily WhatsApp usage

	Outcome: Varies by panel			
	(1)	(2)	(3)	(4)
Panel A: Used Whatsapp				
Treatment	0.056***	0.054***	0.059***	0.047***
	(0.015)	(0.014)	(0.017)	(0.016)
Treatment × NRM supporter			-0.013	0.010
			(0.031)	(0.030)
Treatment + Treatment × NRM supporter			0.046*	0.057**
			(0.026)	(0.025)
Observations	93,124	92,380	93,124	92,380
Clusters (Respondents)	751	745	751	745
Control mean	0.16	0.16	0.16	0.16
Control SD	0.37	0.37	0.37	0.37
Interactive LASSO-selected covariates		✓		✓
Panel B: Number of audited times seen on WhatsApp				
Treatment	0.091***	0.083***	0.096***	0.083***
	(0.024)	(0.024)	(0.029)	(0.028)
Treatment × NRM supporter			-0.023	0.009
			(0.052)	(0.052)
Treatment + Treatment × NRM supporter			0.073*	0.092**
			(0.043)	(0.044)
Observations	93,124	92,380	93,124	92,380
Clusters (Respondents)	751	745	751	745
Control mean	0.27	0.27	0.27	0.27
Control SD	0.68	0.68	0.68	0.68
Interactive LASSO-selected covariates		✓		✓

Notes: Each specification is estimated using OLS, and includes individual and date fixed effects. Even-indexed columns include covariates interacted with treatment period, where covariates are selected by LASSO. To estimate heterogeneous treatment effects, the indicator for NRM supporter is fully interacted with treatment indicator and fixed effects (and LASSO-selected covariates, when relevant). The sum of coefficients reports the treatment effect among NRM supporting participants. We exclude all dates following the conclusion of treatment on September 1, 2021. Standard errors clustered by respondent are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

cares most about citizens ($p = 0.72$), Panel C reports weak evidence that respondents came to feel less warmly about the NRM ($p = 0.15$), and Panel D finds similarly significant estimates that respondents became less willing to vote for the NRM in the future ($p = 0.21$).

However, these average treatment effects mask heterogeneous responses by partisanship. To test our prespecified hypothesis that NRM partisans update more negatively about the ruling party when given greater access to social media, we subset the sample by prior NRM support. Supporting this

Table 2: Field experimental treatment effects on NRM and opposition party support

	Support for NRM				Support for opposition			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Support for NRM and opposition party ICW indexes								
Treatment	-0.060	-0.080*	-0.014	-0.007	0.014	0.053	-0.018	-0.000
	(0.050)	(0.048)	(0.064)	(0.064)	(0.051)	(0.049)	(0.065)	(0.065)
Treatment × NRM supporter			-0.262**	-0.269**			0.207	0.197
			(0.113)	(0.119)			(0.126)	(0.130)
Treatment + Treatment × NRM supporter			-0.276***	-0.276***			0.189*	0.197*
			(0.093)	(0.101)			(0.108)	(0.112)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,387
Control mean	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00
Control SD	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO-selected covariates		✓		✓		✓		✓
Panel B: Which party cares most about people like the respondent								
Treatment	-0.008	-0.022	0.005	0.007	0.026	0.027	0.013	0.010
	(0.022)	(0.021)	(0.030)	(0.031)	(0.021)	(0.020)	(0.029)	(0.029)
Treatment × NRM supporter			-0.085*	-0.076			0.077	0.071
			(0.050)	(0.052)			(0.047)	(0.048)
Treatment + Treatment × NRM supporter			-0.080**	-0.069			0.090**	0.081**
			(0.040)	(0.042)			(0.037)	(0.039)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,387
Control mean	0.72	0.72	0.72	0.72	0.24	0.24	0.24	0.24
Control SD	0.45	0.45	0.45	0.45	0.43	0.43	0.43	0.43
LASSO-selected covariates		✓		✓		✓		✓
Panel C: Feeling thermometer (0-very cold – 10-very warm)								
Treatment	-0.200	-0.253*	-0.110	-0.133	-0.068	-0.091	-0.255	-0.204
	(0.140)	(0.136)	(0.180)	(0.186)	(0.133)	(0.130)	(0.166)	(0.170)
Treatment × NRM supporter			-0.634*	-0.755**			0.599*	0.463
			(0.343)	(0.379)			(0.327)	(0.349)
Treatment + Treatment × NRM supporter			-0.744**	-0.888***			0.344	0.259
			(0.292)	(0.331)			(0.281)	(0.305)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389
Control mean	5.96	5.96	5.96	5.96	5.41	5.41	5.41	5.41
Control SD	2.79	2.79	2.79	2.79	2.64	2.64	2.64	2.64
LASSO-selected covariates		✓		✓		✓		✓
Panel D: Openness to voting for party (1-not at all – 5-very open)								
Treatment	-0.091	-0.100	-0.031	-0.040	-0.002	0.060	0.010	0.030
	(0.073)	(0.071)	(0.095)	(0.095)	(0.071)	(0.069)	(0.091)	(0.091)
Treatment × NRM supporter			-0.227	-0.249			0.134	0.144
			(0.165)	(0.175)			(0.190)	(0.187)
Treatment + Treatment × NRM supporter			-0.258*	-0.289**			0.144	0.174
			(0.135)	(0.147)			(0.166)	(0.164)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389
Control mean	3.35	3.35	3.35	3.35	3.08	3.08	3.08	3.08
Control SD	1.44	1.44	1.44	1.44	1.43	1.43	1.43	1.43
LASSO-selected covariates		✓		✓		✓		✓

Notes: Each specification is estimated using OLS, and adjusts for the baseline and midline pre-treatment outcome and block and wave 3 enumerator fixed effects. Even-indexed columns add LASSO-selected covariates. To estimate heterogeneous treatment effects, the indicator for NRM supporter is fully interacted with treatment indicator and fixed effects (and LASSO-selected covariates, when relevant). Lower-order terms are omitted to save space. The sum of coefficients reports the treatment effect among NRM supporting participants. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

expectation, the conditional average treatment effect captured by the sum of coefficients at the foot of column (3) shows that the modest overall reductions in aggregated attitudes towards the NRM are driven by initial NRM supporters becoming 0.28 standard deviations less favorable ($p < 0.01$). Specifically, NRM supporters became 8 percentage points less likely to believe the NRM cares most ($p < 0.05$), 0.74 units less warm toward the NRM on our 0-10 thermometer ($p < 0.05$), and 0.26 units less open to voting for the NRM in the future on our five-point scale ($p < 0.1$). In turn, NRM supporters became more favorable towards opposition parties ($p < 0.1$), which is largely driven by coming to believe that opposition parties care more about citizens than the NRM.

In contrast, non-NRM supporters updated less in response to treatment. Consistent with having already internalized critical perspectives about the NRM, the base treatment coefficients in columns (3) and (7) show that these participants logged negligible and statistically insignificant reductions in their support for *both* ruling and opposition parties. The interaction coefficients confirm that NRM supporters updated significantly more negatively about the NRM than non-NRM supporters across our various outcomes.

Our finding that facilitating access to social reduced support for the ruling party among prior NRM supporters survives various sensitivity analyses. First, even columns in Table 2 show that our estimates are robust to adjusting for the covariates retained by Belloni, Chernozhukov and Hansen's (2014) double-selection LASSO algorithm, which renders the negative average treatment effect weakly statistically significant ($p = 0.10$). Second, Table A10 shows that our estimates are robust to instead constructing our outcome indexes by taking the difference between outcomes for NRM and opposition parties, using a z-score, or taking the first principal component. Third, Table A11 shows that respondents are not affected by the number of treated individuals in their trading center, suggesting that our results do not reflect interference across participants. Finally, suggesting the absence of demand effects, Table A12 finds no evidence that the intervention—whether in the full sample or by partisanship—affected either participants' perceptions of who was responsible for the survey or their perceptions of the study's purpose.

5.4 Mechanisms connecting social media access with incumbent support

We interpret the heterogeneous effects of subsidizing social media access as NRM supporters, who had initially internalized less regime-critical content than non-NRM supporters, updating negatively about the ruling party upon exposure to more critical content on social media. Several analyses substantiate this mechanism.

First, in line with this belief updating logic, we find that greater treatment intensity induced larger drops in NRM support. Table A13 shows that decreased NRM support among initial supporters was somewhat more pronounced among VPN users who could circumvent the ongoing

Facebook ban to access further critical content. Moreover, Figure A9 shows that treatment effects are larger among NRM-supporting respondents predicted—by our causal forest—to have increased their social media use the most in response to treatment.

Second, our evidence further suggests that greater social media access reduced NRM support by exposing NRM supporters to persuasive regime-critical perspectives rather than by increasing their knowledge of current events. Additional pre-registered outcomes in Figure A8 provide little systematic evidence that treatment led respondents to learn about specific political events that took place during the treatment period, nor to more correctly answer factual questions about politics.[18] Rather, respondents—and particularly NRM supporters—reduced their approval of government performance and became more likely to distrust information from the government.

Third, changes in NRM support were concentrated among the NRM supporters with the greatest prior levels of faith in the government. Panel A of Table A14 shows that NRM supporters who perceived the central government as performing well before the intervention reduced their support for the government significantly more than those who did not. Furthermore, in line with their more limited prior exposure to regime-critical information, the same was true among those NRM supporters who initially reported never having seen information helping them to understand the perspective of opposition parties. By contrast, Panel B shows that neither variation in NRM supporters' prior knowledge of recent political events (measured by the share of five recent political events they correctly identify as having occurred) nor their factual political knowledge (measured by the share of three subnational political leaders they correctly name) moderated an individual's response to treatment.

Together, our experimental results suggest that enabling ruling party supporters to consume more social media moderated their political views—away from the NRM—in Uganda's electoral autocracy. This partial equilibrium finding aligns with the hopes of some that social media might buttress opposition movements in authoritarian regimes, by exposing individuals to social media content that is less regime-friendly than government-dominated traditional media markets, rather than facilitating the echo chambers often found in the Global North.

6 Access to social media around the election-time ban

Our partial equilibrium finding that social media access reduced prior supporters' favorability towards the ruling party may concern incumbents. Indeed, two days before Uganda's 2021 presidential election, the government introduced a month-long ban of social media platforms and then

[18] Additional analyses show that respondents' perceptions of Ugandan state capacity, the quality of its democracy, and local accountability were not moved by treatment.

a five-day full internet blackout around election day. While the social media ban limited access among citizens without VPNs, and may have particularly discouraged content production by regime opponents afraid of unevenly-applied sanctions, it also likely alienated citizens relying on social media for political news, social connection, or economic opportunities. We next investigate how content shifts and backlash endogenous to government policy decisions shape the more general equilibrium effects of restricting access to social media on support for the ruling party.

6.1 Difference-in-differences design

To estimate the effects of differential access to social media during the election-time social media ban, we leverage a difference-in-differences design comparing individuals who already used VPNs prior to the ban—who could easily circumvent the social media ban—with those who did not. Our baseline specification compares changes in NRM support between survey waves 1 and 2 across the 57% of individuals who reported using a VPN on one or more days in the week preceding the wave 1 survey and the remaining individuals who reported no usage. Figure A11 reports the distribution of prior VPN usage. Since non-VPN users could start using VPNs, our design likely underestimates the effect of a fully enforced ban.

VPN users, unsurprisingly, differ from non-VPN users. Table A18 shows that wave 1 VPN users are younger, more likely to use social media, and more favorable towards the opposition. However, these VPN users are similar in various other ways, including gender, religion, education, and self-assessed living conditions; this may reflect the cost of the social media tax roughly netting out with the greater data costs of using a VPN. Conditioning on location and age, by introducing trading center and age fixed effects respectively, largely reduces differences between VPN and non-VPN users to statistical insignificance. To ensure that (observed or unobserved) baseline differences across users are not driving our findings, we exploit within-individual variation over time; to mitigate against time-varying effects of differences, we allow covariate adjustment to vary across time.

To estimate the effect of being more likely to retain access to social media during the government's ban, we estimate difference-in-differences regressions of the form:

$$Y_{ict} = \tau(VPN_i \times Post\ election_t) + \mu_i + \eta_t + \varepsilon_{ict}, \tag{2}$$

where Y_{ict} denotes an outcome for individual i located in trading center c at time t (whether a survey wave or a measure of social media activity), $Post\ election_t$ indicates the period after social media was blocked by the government, and VPN_i indicates prior VPN users. We include individual fixed effects, μ_i, and time fixed effects, η_t, to absorb time-invariant differences across individuals and common period shocks. The former abstracts from baseline differences across respondents who differ in their use of VPNs, while the latter absorbs period-specific factors that influenced all respondents similarly. Standard errors are clustered at the trading center level to reflect community-level differences in VPN usage.

The coefficient τ captures the average effect of already using a VPN during the election-time social media ban—and its aftermath—relative to not previously using a VPN under a parallel trends assumption. This assumption requires that VPN and non-VPN users would have followed similar trends in social media usage and government support outcomes in the absence of the ban. Figure 6 supports this assumption by showing parallel pre-trends in WhatsApp usage across these groups prior to the ban. Parallel trends are harder to substantiate for political outcomes because we only surveyed respondents once before the ban. Nonetheless, Figure A12 supports this assumption by showing that our measures of party support *by survey enumeration date* followed similar trends across VPN and non-VPN users prior to the end of the wave 1 survey. Our robustness checks below further include interactive fixed effects to exploit only variation within various groups—by trading center, age, political engagement and knowledge, and prior political disposition—that could have experienced non-parallel trends.[19]

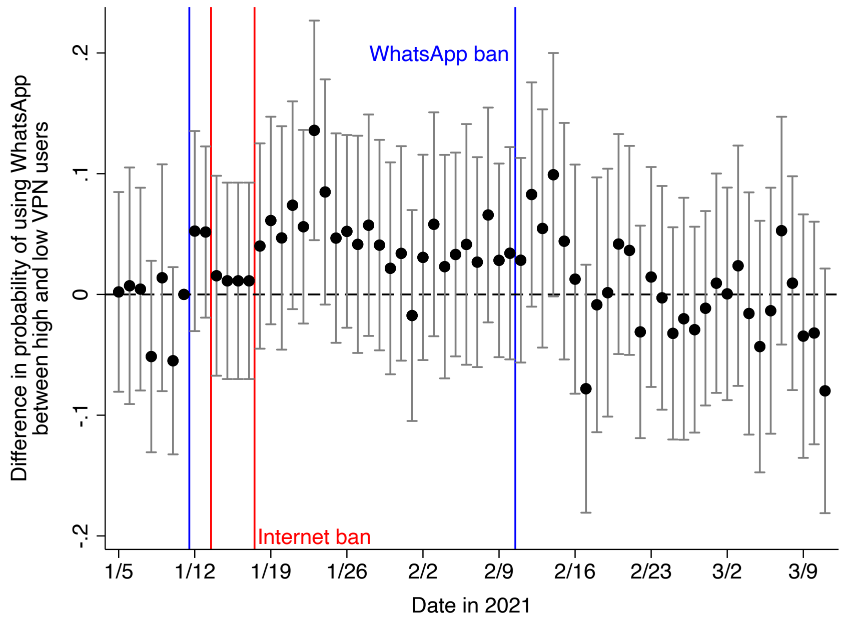
6.2 Effects of variation in exposure to the ban on social media usage

We first confirm that prior VPN users were more likely to access social media during the ban, using the WhatsApp audit data described above. Figure 6 plots our difference-in-differences estimates by day, relative to the day before the social media ban was imposed. While the daily estimates are noisy, the average differences over time are clear: VPN users became more likely to use WhatsApp on a given day during the social media ban—except during the internet blackout—and quickly returned to pre-ban differentials once all platforms (except Facebook) were reinstated in mid-February. Relatively greater social media use among initial VPN users, net of non-VPN users starting to acquire VPNs several weeks after the ban's imposition, is thus concentrated during the month after Uganda's presidential election.

Our regression analyses in Table 3 test this relationship by pooling ban versus non-ban days. Panel A shows that regular VPN users became 5.4 percentage points more likely to use WhatsApp on a given day during the social media ban ($p < 0.01$), relative to a baseline level of 0.31 among non-VPN users. The magnitude of this relative difference in usage is similar to our experimental intervention, although—pointing to its political salience—overall rates of WhatsApp usage were higher during the electoral period in spite of the ban. The non-zero rate among non-VPNs users

[19]The parallel trends assumption could also be violated by differential attrition, but Table A19 shows that VPN and non-VPN users dropped out of the wave 2 survey at statistically indistinguishable rates (about 15%).

Figure 6: Differences in daily use of WhatsApp, by prior VPN use



Notes: Estimates are from equation (2), where the outcome is daily use of WhatsApp. The baseline category is the day before the WhatsApp ban was imposed. All bars indicate 95% confidence intervals.

indicates that a significant share of these individuals started to more regularly use VPNs during the social media ban, in line with prior evidence of citizens' resilience against censorship (e.g. Chang et al. 2022; Roberts 2020). Figure A14 reports similar results when instead defining VPN users by using a VPN more than one day in the week preceding survey wave 1.

Like our experimental analysis, we examine heterogeneity in the difference-in-differences treatment effect by estimating a causal forest.[20] Figure A10 shows that the respondents most likely to increase their WhatsApp usage due to VPN availability were relatively better educated and used more social media at baseline, while being relatively less supportive of the government. Table A20 also provides tentative evidence of positive effects on usage among VPN-using NRM supporters as well as non-NRM supporters.[21]

[20]We follow the same broad approach, using the difference in the mean share of days on which a respondent used WhatsApp during the ban relative to before as our outcome, and with all predictive variables defined pre-ban. Based on Table A18, we residualize our indicator for VPN usage using age and trading center fixed effects to obtain a treatment variable more plausibly unconfounded in the cross-section.

[21]In contrast with the field experiment, disaggregating by partisanship poses inferential problems because regime opposition plausibly *causes* their uptake of VPNs to begin with; this difficulty is most pronounced for our political outcomes where baseline outcomes are used to construct the partisanship moderator. Disaggregating on this post-treatment basis creates a biased, and difficult to interpret, quantity regardless of outcome.

Table 3: Differential effects of prior VPN use on daily WhatsApp usage during the social media ban

	Outcome: Varies by panel		
	(1)	(2)	(3)
Panel A: Used Whatsapp			
VPN × WhatsApp ban	0.054***	0.047***	0.046**
	(0.014)	(0.017)	(0.018)
Control mean	0.31	0.30	0.30
Control SD	0.46	0.46	0.46
Panel B: Number of audited times seen on WhatsApp			
VPN × WhatsApp ban	0.084***	0.077**	0.076**
	(0.028)	(0.034)	(0.033)
Observations	112,943	111,655	111,655
Clusters (TCs)	125	116	116
Control mean	0.54	0.53	0.53
Control SD	0.95	0.95	0.95
Interactive fixed effects		TC	TC & Age

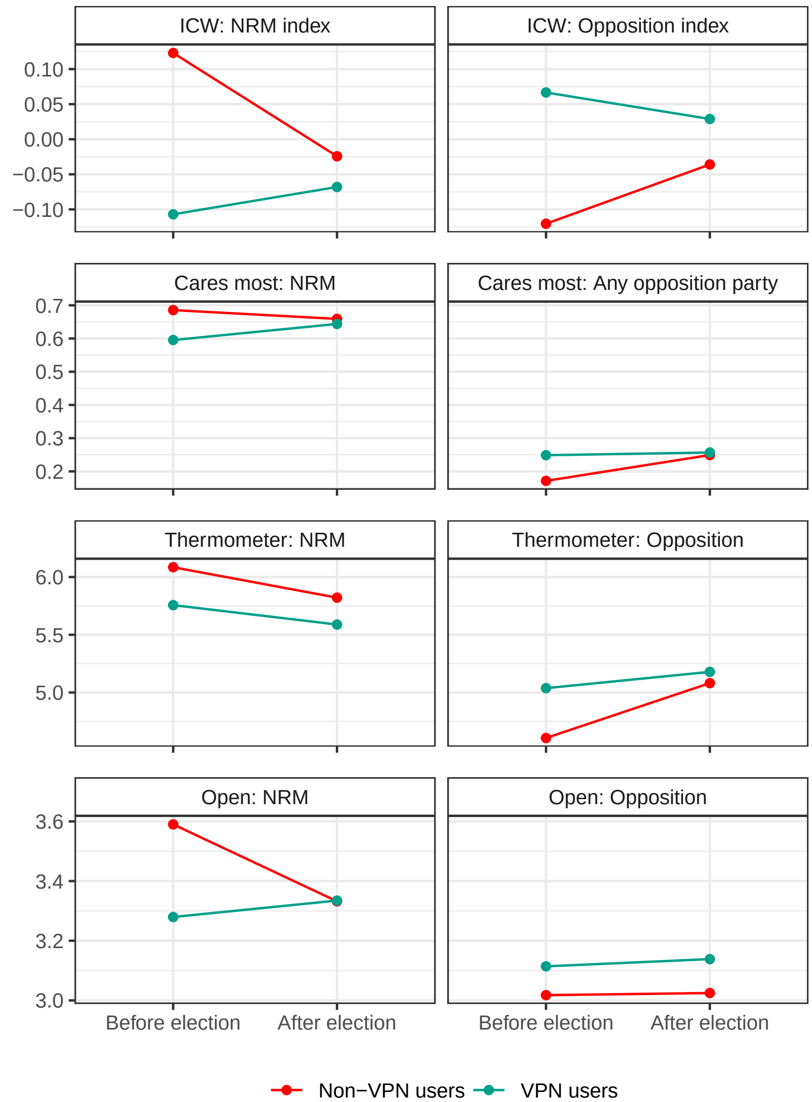
Notes: Each specification is estimated using OLS, and includes individual and period fixed effects. Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6.3 Increased relative support for the NRM governing party

We next examine changes in support for the ruling and opposition parties using the same outcome measures as the field experiment. Figure 7 shows that prior VPN users were less likely to believe that the NRM cares most about Ugandans' welfare, less warm about the NRM, and less open to voting for an NRM candidate in the future in our pre-election survey. By wave 2, after the social media ban had been imposed and withdrawn (except for Facebook) and Museveni had been inaugurated, prior VPN users viewed the NRM slightly more positively, in absolute terms, while non-VPN users viewed the NRM more negatively. In contrast, non-VPN users became more favorable toward opposition parties. Non-VPN users switching their support from NRM to opposition parties might suggest that a backlash effect dominates, but absolute support levels could also reflect common period shocks—for example, NRM support may universally subside after elections because citizens are not being mobilized by campaigns. Consequently, our difference-in-differences design most credibly reveals that VPN users became more favorable toward the ruling NRM party *relative* to non-VPN users.

We formally test for relative changes in support by prior VPN use in Table 4 by estimating equation (2), including trading center × period fixed effects in even-numbered columns. The results

Figure 7: Changes in NRM support between pre-election wave 1 and post-election wave 2 surveys, by prior VPN use



Note: Figure plots the raw mean of our various outcome measures in the wave 1 and 2 surveys across respondents by reported VPN use in the wave 1 survey.

show that the gap between VPN users and non-VPN users significantly shrunk after the ban, with the NRM support index outcome in columns (1) and (2) of Panel A increasing from wave 1 to wave 2 by almost 0.2 standard deviations among VPN users relative to non-VPN users ($p < 0.05$). This narrowing is observed across each outcome: relative to non-VPN users, VPN users became nearly eight percentage points more likely to believe that the NRM cares most about Ugandans' welfare ($p < 0.05$), slightly—but not significantly—warmer toward the NRM ($p = 0.63$), and 0.3 units more open to voting for the NRM in the future ($p < 0.01$). By contrast, panel A shows that the index capturing support for opposition parties decreased by slightly over 0.1 standard deviations.

This is driven by relative reductions in their perceptions of opposition parties caring most about Ugandans' welfare as well as more negative feelings towards opposition parties overall. Noting the econometric difficulties in interpretation described above, heterogeneous effects by partisanship in Figure A13 and Table A21 suggest larger effects among non-NRM supporters.

Together, and in contrast with the partial equilibrium effects of subsidizing access during normal times, the difference-in-differences results suggest that the net relative effect of an individual having greater access to social media—via prior VPN use—during the election-time ban was to *increase* NRM support.

Our estimates are robust to alternative specifications and interpretations. First, we address potential parallel trend violations and compound treatment concerns. As well as exploiting only variation in respondents' VPN use within trading centers (by period) in Table 4's even columns, Table A22 reports similar results after adjusting for interactions between period and baseline levels of respondent age, prior political news consumption, political knowledge, and prior support for the NRM. Table A23 shows that our findings are not sensitive to applying entropy balancing over the same covariates. These tests suggest our estimates are specifically driven by access to social media rather than different trends in NRM support among young people, politically-engaged citizens, anti-NRM respondents, or people in particular areas.

Second, our findings are robust across operationalizations of VPN use. Figure A14 reports similar results, across each outcome, when defining VPN users as respondents who used a VPN more than one day a week before wave 1 enumeration.

Finally, an alternative interpretation is that self-reported beliefs reflect socially-desirable responses. In particular, violators of the social media ban may fear punishment and inaccurately profess greater support for the NRM to compensate, which would upwardly bias our estimates. We find little evidence consistent with this possibility: Table A24 shows that VPN users became no more likely to believe the survey firm had been sent by the government or the NRM, while Table A22 shows that our results are robust to adjusting for the interaction between survey period and an indicator for respondents who thought, at wave 1, that the government or NRM was responsible for the study.

6.4 Mechanisms connecting the social media ban with political attitudes

The effects of access to social media notably differ between the randomized subsidy and election-time social media ban. Despite changing social media usage by similar amounts, the former intervention found greater access reduced support for the NRM whereas the latter found the reverse. One possible explanation is that the social media ban coincided with a period of unusually favorable news events for the NRM, enabling social media's more credible content to persuade its

Table 4: Differential effects of prior VPN use on support for the NRM after the social media ban

	Support for NRM		Support for Opposition	
	(1)	(2)	(3)	(4)
Panel A: Support for NRM and opposition party ICW indexes				
VPN × Post election	0.184**	0.196***	-0.114*	-0.156**
	(0.072)	(0.074)	(0.069)	(0.070)
Observations	2,620	2,612	2,620	2,612
Control mean	-0.05	-0.06	0.06	0.07
Control SD	1.00	1.00	0.98	0.98
Trading center × Post election fixed effects		✓		✓
Panel B: Which party cares most about people like the respondent				
VPN × Post election	0.075**	0.078**	-0.070**	-0.071**
	(0.034)	(0.035)	(0.032)	(0.033)
Observations	2,620	2,612	2,620	2,612
Control mean	0.64	0.64	0.23	0.24
Control SD	0.48	0.48	0.42	0.42
Trading center × Post election fixed effects		✓		✓
Panel C: Feeling thermometer (0-very cold – 10-very warm)				
VPN × Post election	0.095	0.285	-0.334*	-0.438**
	(0.196)	(0.208)	(0.172)	(0.181)
Observations	2,620	2,612	2,620	2,612
Control mean	5.79	5.79	4.99	5.00
Control SD	2.72	2.72	2.52	2.52
Trading center × Post election fixed effects		✓		✓
Panel D: Openness to voting for party (1-not at all – 5-very open)				
VPN × Post election	0.313***	0.258**	0.017	-0.044
	(0.112)	(0.121)	(0.116)	(0.122)
Observations	2,620	2,612	2,620	2,612
Control mean	3.37	3.37	3.08	3.09
Control SD	1.43	1.43	1.45	1.45
Trading center × Post election fixed effects		✓		✓

Notes: Each specification is estimated using OLS, and includes individual and period fixed effects. Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

consumers (Heo and Zerbini 2024). But, the interventions also fundamentally differ in their equilibrium effects. Whereas the field experiment subsidized access for a small number of individuals in isolation, the government's social media ban restricted content production and consumption for

Table 5: Differential effects of prior VPN use on potential mediators of support for the NRM after the social media ban

	Democracy with major problems		Follow national government officials		Follow opposition politicians		Central government performance assessment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VPN × Post election	-0.029	-0.046	-0.041	-0.022	-0.075**	-0.086**	0.051	0.121
	(0.031)	(0.032)	(0.037)	(0.040)	(0.033)	(0.035)	(0.081)	(0.092)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
R^2	0.56	0.61	0.57	0.62	0.60	0.64	0.55	0.61
Control outcome mean	0.55	0.55	0.41	0.41	0.40	0.40	3.28	3.28
Control outcome std. dev.	0.50	0.50	0.49	0.49	0.49	0.49	1.15	1.15
Trading center × Post election fixed effects		✓		✓		✓		✓

Notes: Each specification is estimated using OLS, and includes individual and period fixed effects. Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

millions of citizens simultaneously. The ban may thus have altered the types of people producing content, and what content they produced during the ban, while eliciting backlash from non-VPN users more likely to have lost valued access to social media. While we cannot formally quantify each mechanism's relative contribution to the effect of the ban, we find evidence suggesting that both mechanisms are at play.[22]

6.4.1 News content relating to election integrity

Violent repression of opposition protests and leaders in late 2020 and opposition efforts to coordinate online reporting of electoral malpractice were fairly common topics of posts on Facebook early in the election campaign (see Figure A7). Consequently, citizens may have expected extensive reports of fraud or violence on election day that did not ultimately materialize (the reports, that is). We explore whether relatively favorable content on social media during the government's election-time ban contributed to prior VPN users becoming comparatively more supportive of the NRM by examining both posterior beliefs about the quality of democracy in Uganda and whether increased support was concentrated among the respondents with the lowest expectations of fair elections.

We find suggestive evidence that greater access to social media during the ban led citizens to update relatively favorably about Uganda's democracy. First, columns (1) and (2) of Table 5 show that prior VPN users became 3-5 percentage points less likely to say they believed Uganda is a democracy with major problems or not a democracy at all, relative to the 55% of non-VPN users who believed this. This reduction in skepticism is more sharply estimated with covariate adjustment

[22]Appendix A.6 shows that our results cannot be reconciled on the basis of differences in *whose* social media access changed most in the two designs—whether in terms of estimating samples or in the characteristics of the effective "compliers" in the two designs.

Table 6: Heterogeneity in difference-in-differences effects on NRM support, by respondent prior beliefs

	NRM support index (ICW)					
	(1)	(2)	(3)	(4)	(5)	(6)
VPN × Post election	0.110	0.125	0.074	0.049	0.071	0.077
	(0.078)	(0.081)	(0.097)	(0.098)	(0.092)	(0.097)
VPN × Post election × Uganda a flawed democracy prior	0.383**	0.391**				
	(0.191)	(0.191)				
VPN × Post election × Followed opposition politicians			0.195	0.272**		
			(0.132)	(0.134)		
VPN × Post election × Non-good incumbent performance prior					0.215*	0.233*
					(0.125)	(0.140)
VPN × Post election + VPN × Post election × covariate	0.492***	0.517***	0.270***	0.321***	0.286***	0.311***
	(0.170)	(0.169)	(0.100)	(0.103)	(0.100)	(0.109)
Observations	2,620	2,612	2,620	2,612	2,620	2,612
R^2	0.63	0.68	0.62	0.66	0.62	0.67
Control outcome mean	0.00	-0.00	0.00	-0.00	0.00	-0.00
Control outcome std. dev.	1.00	1.00	1.00	1.00	1.00	1.00
Trading center × Post election fixed effects		✓		✓		✓

Notes: Each specification is estimated using OLS, and includes individual and period fixed effects. Standard errors clustered by trading center are in parentheses. Lower-order interaction terms are omitted to save space. The sum of coefficients reports the difference-in-differences estimate when each binary covariate is equal to 1. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

in column (2) ($p = 0.14$). Second, and more tellingly, increased NRM support among prior VPN users after the election is driven by respondents for whom violations of democratic norms were likely to have been less bad than expected. The first two columns of panel A in Table 6 show that increased NRM support is four times larger among prior VPN users who believed Uganda was a democracy with major problems, or not a democracy at all, at baseline. These findings suggest that favorable updating from low expectations about the integrity of the elections—likely to be most pronounced among non-NRM supporters—may help explain increased support for the NRM among VPN users relative to non-VPN users, who may have assumed reports of electoral malfeasance would be widespread on social media.

6.4.2 Changes in social media content around the ban

While prominent news events—or lack thereof—might have become more favorable towards the NRM during the social media ban, the ban also altered the slant of political content encountered on social media more generally. Likely by preventing non-VPN users from using social media and fear of differential sanctions for violating the ban against the NRM's online opponents, Section 3.4 showed that both the share of Facebook content that was critical of NRM, and the share of posts by opposition-leaning accounts, dropped immediately after the ban was imposed. These changes are driven by pages which had been most critical of the government prior to the ban. Individuals retaining access to social media during the ban were thus exposed to content which was *relatively* more pro-government, and less political in general, than during the pre-election period.

Our analyses indicate that VPN users indeed consumed different content and were somewhat swayed by it. Regarding social media content consumption, columns (3) and (4) of Table 5 show that VPN users became 8 percentage points relatively less likely to follow opposition—but not government—politicians on social media than non-VPN users after the social media ban. Turning to posterior beliefs, columns (5) and (6) further report that positive appraisals of central government performance modestly and statistically insignificantly increased on a five-point scale ranging from very bad (1) to very good (5) among prior VPN users relative to non-VPN users. Finally, the heterogeneous effects in Table 6 document a significantly larger relative increase in support among initial VPN users who followed opposition politicians or had low perceptions of central government performance at baseline, which are again likely to be concentrated among prior non-NRM supporters. These results suggest that changes in the slant of online content during the social media ban may also have contributed to citizens with greater access to social media updating favorably toward the NRM.

6.4.3 Sanctioning of government by non-VPN users

A third potential mechanism helping to explain why the effect of government policies limiting access to social media may differ from our partial equilibrium experimental results is that non-VPN users were more likely to sanction government censorship. While the government's ban limited everyone's access to social media, its impact on the livelihoods of people not already using VPNs would have been greater. Under this interpretation, the positive difference-in-differences estimate may reflect non-VPN users becoming relatively less supportive of the NRM, rather than VPN users becoming more supportive of the NRM.

We evaluate this possibility by examining whether lost support was concentrated among the respondents most adversely affected by the social media ban. Our wave 2 survey asked respondents if the social media and internet restrictions interfered their business/job (40% of respondents answered affirmatively), affected their ability to purchase goods and services (10%), reduced their ability to talk to friends and family (78%), reduced their ability to find reliable news (58%), and affected their ability to consume online entertainment content (18%). To investigate the effects of suffering from the social media ban in these ways, we estimate the following difference-in-differences regression:

$$Y_{ict} = \tau(Censorship_i \times Post\ election_t) + \mu_i + \eta_t + \varepsilon_{ict}, \tag{3}$$

where $Censorship_i$ captures a particular cost of censorship that a respondent recalled incurring.

Our findings in Table 7 provide evidence of a backlash effect. Columns (1)-(4) show that

Table 7: Differential effects of censorship experience on support for the NRM after the social media ban

	NRM support index (ICW)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ban reduced access to reliable news × Post election	-0.175***	-0.185**								
	(0.065)	(0.071)								
Ban reduced social interactions × Post election			-0.142	-0.148						
			(0.090)	(0.097)						
Ban affected access to entertainment × Post election					-0.063	-0.054				
					(0.086)	(0.089)				
Ban affected purchase of goods/services × Post election							-0.087	-0.074		
							(0.131)	(0.140)		
Ban interfered with business/job × Post election									0.142*	0.074
									(0.074)	(0.075)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
R^2	0.62	0.66	0.62	0.66	0.62	0.66	0.62	0.66	0.62	0.66
Ban variable mean	0.58	0.58	0.78	0.78	0.18	0.18	0.10	0.10	0.40	0.40
Trading center × Post election fixed effects		✓		✓		✓		✓		✓

Notes: Each specification is estimated using OLS, and includes individual and period fixed effects. Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

respondents who reported the social media ban reduced their ability to find reliable news and talk with friends and family became significantly less supportive of the NRM, by around 0.15 standard deviations, relative to respondents who did not report such consequences. We similarly observe negative standardized effects among respondents who reported losing access to entertainment and purchasing capacity, although these estimates are not statistically significant. Respondents who said the social media ban affected their livelihood became somewhat more supportive, but this does not hold when exploiting only within-trading center variation in column (10). The results thus point to a net sanctioning effect driven by citizens who lost access to reliable news and connections during the ban. Table A25 suggests this effect operates alongside the content effects induced by prior VPN use.

7 Conclusion

This paper investigated the effects of social media access on incumbent support in Uganda, where traditional media is tightly controlled. Our experimental "partial equilibrium" analysis shows that access to broadly regime-critical social media content reduces support for the ruling incumbent among prior supporters. However, opponents' access to such a megaphone can motivate authoritarian governments, which face few institutional constraints, to restrict access. Leveraging a difference-in-difference design during a partial social media ban to illuminate the trade-offs shaping such decisions, we find that a ban which restricted social media access *en masse* reduced the incidence of content on social media that is critical of the ruling party, in addition to reducing exposure to it, but also induced backlash among citizens who lost access.

These findings highlight the regime-specific effects of social media. Whereas research from established democracies largely points to limited counter-attitudinal effects of social media, especially where users select the content they consume, our evidence from Uganda shows that social media has the potential to serve as a democratizing force by reducing support for elected autocrats. The difference appears to reflect lower barriers to entry in electoral autocracies for producing online content that criticizes government performance, relative to traditional media, as well as social media serving as less of an echo chamber in these often less-partisan settings. It thus makes sense that opposition groups with limited institutional voice, like Bobi Wine's NUP movement, gravitate toward social media, especially when age—a key correlate of social media use—is a salient political cleavage.

The capacity of governments to censor social media is another distinctive feature of autocracies. Although we cannot assess whether Uganda's social media restrictions were successful overall, given the plethora of possible effects and government objectives, our results reveal that governments with limited capacity to comprehensively monitor, selectively remove, or "outcompete" content (Roberts 2020) must evaluate trade-offs between restricting access to content, altering what content is produced, and encountering backlash. This backlash constraint is alleviated for high-capacity autocratic regimes, which are able to identify and target specific pieces of content and prevent workarounds, without necessarily alienating users by significantly diminishing their online experience. As technology develops and governments like China export digital censorship tools to allied regimes, the costs of social media censorship may diminish. This may ultimately undercut social media's democratizing potential by increasing the subtlety, rather than the quantity, of censorship. Our analysis of a context where subtler instruments or censorship are not yet commonplace, especially outside of key moments like elections, provides a starker setting that helps to illuminate the trade-offs governing an autocrat's censorship dilemma.

A limitation of our analysis is that we cannot establish exactly what content individuals consumed on social media during our study periods. Consequently, a key outstanding question is the extent to which social media users are persuaded by different types of content. One possibility, for which we find suggestive evidence, is that citizens learn about government performance. But others include exposure to alternative worldviews and values, novel perspectives on events, or horizontal persuasion by peers. Parsing between these is likely to require different research designs controlling content exposure. In contrast, we identify overall effects of social media when citizens are free to consume as they choose. This overall effect represents a critical question both for those raising alarm bells about the rise of social media and, particularly in electoral autocracies, the governments feeling threatened by it. Ultimately, both empirical components of our study point to substantial political effects of social media. In contrast with deactivation studies that almost entirely end social media use (e.g. Allcott et al. 2020, 2024; Ventura et al. 2023), our experimental and observational studies induced moderate changes across users in social media consumption. Nevertheless, we detect significant effects on incumbent support among individuals who are already marginal social media users. Although this suggests particularly large political ramifications in electoral autocracies, further research is needed to generalize these effects across countries and examine the consequences of using social media for the first time.

References

A4AI. 2019. Mobile Broadband Pricing Data. Technical report.
URL: *https://a4ai.org/extra/mobile_broadband_pricing_gnicm-2019Q2*

Afrobarometer. 2019. “Afrobarometer Data, Uganda, Round 8, 2019.”.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer and Matthew Gentzkow. 2020. “The Welfare Effects of Social Media.” *American Economic Review* 110(3):629–76.

Allcott, Hunt, Matthew Gentzkow, Winter Mason, Arjun Wilkins, Pablo Barberá, Taylor Brown, Juan Carlos Cisneros, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon et al. 2024. “The effects of Facebook and Instagram on the 2020 election: A deactivation experiment.” *Proceedings of the National Academy of Sciences* 121(21):e2321584121.

Amnesty International. 2018. Uganda: Scrap social media tax curtailing freedom of expression. Technical report.
URL: *https://www.amnesty.org/en/latest/news/2018/07/uganda-scrap-social-media-tax-curtailing-freedom-of-expression/*

Amnesty International. 2021. Uganda: Museveni’s latest government must reverse decline on human rights. Technical report.
URL: *https://www.amnesty.org/en/latest/news/2021/05/uganda-musevenis-latest-government-must-reverse-decline-on-human-rights/*

Anderson, Michael L. 2008. “Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of the American Statistical Association* 103(484):1481–1495.

Arias, Eric, Horacio Larreguy, John Marshall and Pablo Querubín. 2022. “Priors rule: When do malfeasance revelations help or hurt incumbent parties?” *Journal of the European Economic Association* 20(4):1433–1477.

Aridor, Guy, Rafael Jiménez-Durán, Ro’ee Levy and Lena Song. forthcoming. “The Economics of Social Media.” *Journal of Economic Literature* .

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P Bumpus, Haohan Chen, M.B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout and Alexander Volfovsky. 2018. “Exposure to opposing views on social media can increase political polarization.” *Proceedings of the National Academy of Sciences* 115(37):9216–9221.

Banerjee, Abhijit V., Selvan Kumar, Rohini Pande and Felix Su. 2011. “Do Informed Voters Make Better Choices? Experimental Evidence from Urban India.” Working paper.

Barbera, Salvador and Matthew O. Jackson. 2020. “A Model of Protests, Revolution, and Information.” *Quarterly Journal of Political Science* 15(3):297–335.

Belloni, Alexandre, Victor Chernozhukov and Christian Hansen. 2014. "Inference on Treatment Effects After Selection Among High-dimensional Controls." *Review of Economic Studies* 81(2):608–650.

Besley, Timothy and Andrea Prat. 2006. "Handcuffs for the Grabbing Hand? Media Capture and Government Accountability." *American Economic Review* 96(3):720–736.

Bhandari, Abhit, Horacio Larreguy and John Marshall. 2023. "Able and Mostly Willing: An Empirical Anatomy of Information's Effect on Voter-Driven Accountability in Senegal." *American Journal of Political Science* 67(4):1040–1066.

Blackwell, Matthew and Michael P. Olson. 2022. "Reducing Model Misspecification and Bias in the Estimation of Interactions." *Political Analysis* 30(4):495–514.

Boxell, Levi and Zachary Steinert-Threlkeld. 2022. "Taxing dissent: The impact of a social media tax in Uganda." *World Development* 158:105950.

Braghieri, Luca, Ro'ee Levy and Alexey Makarin. 2022. "Social media and mental health." *American Economic Review* 112(11):3660–3693.

Brierley, Sarah, Eric Kramon and George Kwaku Ofosu. 2020. "The moderating effect of debates on political attitudes." *American Journal of Political Science* 64(1):19–37.

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov and Maria Petrova. 2024. "Social Media and Xenophobia: Theory and Evidence from Russia.".

Callen, Michael, Jonathan L. Weigel and Noam Yuchtman. forthcoming. "Experiments about institutions." *Annual Review of Economics* .

Carlson, Elizabeth. 2016. "Finding Partisanship Where we Least Expect It: Evidence of Partisan Bias in a New African Democracy." *Political Behavior* 38(1):129–154.

Chang, Keng-Chi, William R. Hobbs, Margaret E. Roberts and Zachary C. Steinert-Threlkeld. 2022. "COVID-19 increased censorship circumvention and access to sensitive topics in China." *Proceedings of the National Academy of Sciences* 119(4):e2102818119.

Chen, Yuyu and David Y. Yang. 2019. "The Impact of Media Censorship: 1984 or Brave New World?" *American Economic Review* 109(6):2294–2332.

Chwe, Michael Suk-Young. 2000. "Communication and Coordination in Social Networks." *Review of Economic Studies* 67(1):1–16.

Conroy-Krutz, Jeffrey and Devra C. Moehler. 2015. "Moderation from bias: A field experiment on partisan media in a new democracy." *Journal of Politics* 77(2):575–587.

Cornand, Camille and Frank Heinemann. 2008. "Optimal Degree of Public Information Dissemination." *Economic Journal* 118(528):718–742.

Di Tella, Rafael and Ignacio Franceschelli. 2011. "Government Advertising and Media Coverage of Corruption Scandals." *American Economic Journal: Applied Economics* 3(4):119–151.

Diamond, Larry. 2010. "Liberation Technology." *Journal of Democracy* 21(3):69–83.

Donati, Dante. 2023. "Mobile Internet access and political outcomes: Evidence from South Africa." *Journal of Development Economics* 162:103073.

Egorov, Georgy, Sergei Guriev and Konstantin Sonin. 2009. "Why Resource-poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data." *American Political Science Review* 103(4):645–668.

Enikolopov, Ruben, Alexey Makarin and Maria Petrova. 2020. "Social media and protest participation: Evidence from Russia." *Econometrica* 88(4):1479–1514.

Enikolopov, Ruben, Maria Petrova and Ekaterina Zhuravskaya. 2011. "Media and political persuasion: Evidence from Russia." *American Economic Review* 101(7):3253–3285.

Enikolopov, Ruben, Maria Petrova, Gianluca Russo and David Yanagizawa-Drott. 2024. "Socializing Alone: How Online Homophily Has Undermined Social Cohesion in the US." Working paper.

Enríquez, José Ramón, Horacio Larreguy, John Marshall and Alberto Simpser. 2024. "Mass political information on social media: Facebook ads, electorate saturation, and electoral accountability in Mexico." *Journal of the European Economic Association* 22(4):1678–1722.

Freedom House. 2021. "Freedom in the World 2021: Uganda.".
URL: *https://freedomhouse.org/country/uganda/freedom-world/2021*

Fujiwara, Thomas, Karsten Müller and Carlo Schwarz. 2024. "The effect of social media on elections: Evidence from the united states." *Journal of the European Economic Association* 22(3):1495–1539.

Gehlbach, Scott and Konstantin Sonin. 2014. "Government control of the media." *Journal of Public Economics* 118:163–171.

Gläßel, Christian and Katrin Paula. 2020. "Sometimes less is more: Censorship, news falsification, and disapproval in 1989 East Germany." *American Journal of Political Science* 64(3):682–698.

Guess, Andrew M., Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow et al. 2023. "How do social media feed algorithms affect attitudes and behavior in an election campaign?" *Science* 381(6656):398–404.

Guriev, Sergei and Daniel Treisman. 2019. "Informational autocrats." *Journal of Economic Perspectives* 33(4):100–127.

Guriev, Sergei and Daniel Treisman. 2020. "A theory of informational autocracy." *Journal of Public Economics* 186:104158.

Guriev, Sergei and Daniel Treisman. 2022. *Spin Dictators: The Changing Face of Tyranny in the 21st Century*. Princeton University Press.

Guriev, Sergei, Nikita Melnikov and Ekaterina Zhuravskaya. 2021. "3g internet and confidence in government." *Quarterly Journal of Economics* 136(4):2533–2613.

Heo, Kun and Antoine Zerbini. 2024. "Segment and Rule: Modern Censorship in Authoritarian Regimes." Working paper.

Huang, Haifeng, Serra Boranbay-Akan and Ling Huang. 2019. "Media, protest diffusion, and authoritarian resilience." *Political Science Research and Methods* 7(1):23–42.

Kern, Holger Lutz and Jens Hainmueller. 2009. "Opium for the masses: How foreign media can stabilize authoritarian regimes." *Political Analysis* 17(4):377–399.

King, Gary, Jennifer Pan and Margaret E. Roberts. 2013. "How censorship in China allows government criticism but silences collective expression." *American Political Science Review* 107(2):326–343.

King, Gary, Jennifer Pan and Margaret E. Roberts. 2014. "Reverse-engineering censorship in China: Randomized experimentation and participant observation." *Science* 345(6199).

King, Gary, Jennifer Pan and Margaret E. Roberts. 2017. "How the Chinese government fabricates social media posts for strategic distraction, not engaged argument." *American Political Science Review* 111(3):484–501.

Kronick, Dorothy and John Marshall. 2024. "Collateral censorship: Theory and evidence from Venezuela." Working paper.

Kuran, Timur. 1989. "Sparks and prairie fires: A theory of unanticipated political revolution." *Public choice* 61(1):41–74.

Lawson, Chappell and James A. McCann. 2005. "Television news, Mexico's 2000 elections and media effects in emerging democracies." *British Journal of Political Science* 35(1):1–30.

Letsa, Natalie Wenzell. forthcoming. "Partisanship and Political Socialization in Electoral Autocracies." *American Political Science Review* .

Letsa, Natalie Wenzell and Yonatan L Morse. 2023. "Autocratic Legalism, Partisanship, and Popular Legitimation in Authoritarian Cameroon." *Public Opinion Quarterly* 87(4):935–955.

Levitsky, Steven and Lucan Way. 2010. *Competitive Authoritarianism: Hybrid Regimes after the Cold War*. Cambridge: Cambridge University Press.

Levy, Ro'ee. 2021. "Social media, news consumption, and polarization: Evidence from a field experiment." *American Economic Review* 111(3):831–870.

Little, Andrew T. 2023. "Bayesian explanations for persuasion." *Journal of Theoretical Politics* 35(3):147–181.

Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science* 58(2):402–414.

Majumdar, Rajeshwari. forthcoming. “Partisan Identity, Counter-Attitudinal Information, and Selective Criticism in India.” *Political Behavior* .

Manacorda, Marco, Edward Miguel and Andrea Vigorito. 2011. “Government Transfers and Political Support.” *American Economic Journal: Applied Economics* 3(3):1–28.

McMillan, John and Pablo Zoido. 2004. “How to Subvert Democracy: Montesinos in Peru.” *Journal of Economic Perspectives* 18(4):69–92.

Miller, Andrew Cesare. 2022. “#DictatorErdogan: How Social Media Bans Trigger Backlash.” *Political Communication* 39(6):801–825.

Miner, Luke. 2015. “The unintended consequences of internet diffusion: Evidence from Malaysia.” *Journal of Public Economics* 132:66–78.

Morozov, Evgeny. 2012. *The Net Delusion: The Dark Side of Internet Freedom*. PublicAffairs.

Morris, Stephen and Hyun Song Shin. 2002. “Social Value of Public Information.” *American Economic Review* 92(5):1521–1534.

Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo and Ragan Petrie. 2020. “The Economic Effects of Facebook.” *Experimental Economics* 23(2):575–602.

Müller, Karsten and Carlo Schwarz. 2023. “From hashtag to hate crime: Twitter and antiminority sentiment.” *American Economic Journal: Applied Economics* 15(3):270–312.

Namasinga, Florence and Kristin Orgeret. 2020. “Social media in Uganda: revitalising news journalism?” *Media, Culture & Society* 42(3):380–397.

Nyabola, Nanjala. 2018. *Digital democracy, analogue politics: How the Internet era is transforming politics in Kenya*. Bloomsbury Publishing.

Nyhan, Brendan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y. Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery et al. 2023. “Like-minded sources on Facebook are prevalent but not polarizing.” *Nature* 620(7972):137–144.

Peterson, Erik and Ali Kagalwala. 2021. “When unfamiliarity breeds contempt: How partisan selective exposure sustains oppositional media hostility.” *American Political Science Review* 115(2):585–598.

Platas, Melina R. and Pia J. Raffler. 2021. “Closing the gap: Information and mass support in a dominant party regime.” *Journal of Politics* 83(4):1619–1634.

Pollicy. 2020. A Shot in the Dark: The Impact of the Social Media Tax in Uganda on Access, Usage, Income and Productivity. Technical report.
URL: *http://pollicy.org/wp-content/uploads/2020/03/A-Shot-in-the-Dark-The-Impact-of-the-Social-Media-Tax-in-Uganda.pdf*

Prior, Markus. 2007. *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.

Qin, Bei, David Strömberg and Yanhui Wu. 2017. "Why does China allow freer social media? Protests versus surveillance and propaganda." *Journal of Economic Perspectives* 31(1):117–140.

Qin, Bei, David Strömberg and Yanhui Wu. 2018. "Media bias in China." *American Economic Review* 108(9):2442–2476.

Reuter, Ora John and David Szakonyi. 2021. "Electoral Manipulation and Regime Support: Survey Evidence from Russia." *World Politics* 73(2):275–314.

Roberts, Margaret E. 2018. *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton University Press.

Roberts, Margaret E. 2020. "Resilience to online censorship." *Annual Review of Political Science* 23:401–419.

Rosenfeld, Bryn. 2020. *The Autocratic Middle Class: How State Dependency Reduces the Demand for Democracy*. Vol. 11 Princeton University Press.

Shadmehr, Mehdi and Dan Bernhardt. 2015. "State censorship." *American Economic Journal: Microeconomics* 7(2):280–307.

Shadmehr, Mehdi and Dan Bernhardt. 2017. "When Can Citizen Communication Hinder Successful Revolution?" *Quarterly Journal of Political Science* 12(3):301–323.

Simpser, Alberto. 2013. *Why Governments and Parties Manipulate Elections: Theory, Practice, and Implications*. Cambridge University Press.

Steinert-Threlkeld, Zachary C. 2017. "Spontaneous collective action: Peripheral mobilization during the Arab Spring." *American Political Science Review* 111(2):379–403.

Sunstein, Cass. 2018. *# Republic: Divided democracy in the age of social media*. Princeton University Press.

Szeidl, Adam and Ferenc Szucs. 2021. "Media capture through favor exchange." *Econometrica* 89(1):281–310.

Tucker, Joshua, Yannis Theocharis, Margaret Roberts and Pablo Barberá. 2017. "From liberation to turmoil: social media and democracy." *Journal of Democracy* 28(4):46–59.

Uganda Communications Commission. 2021. "Market Performance Report Q1 2021.".
URL: *https://www.ucc.co.ug/wp-content/uploads/2021/07/UCC*$_1$*Q21 − Market − Performance − Rerport_compressed.pdf*

US Department of State. 2020. "Country Reports on Human Rights Practices: Uganda.".
URL: *https://www.state.gov/wp-content/uploads/2021/02/Uganda-2020-Human-Rights-Report.pdf*

Ventura, Tiago, Rajeshwari Majumdar, Jonathan Nagler and Joshua A. Tucker. 2023. "Misinformation Exposure Beyond Traditional Feeds: Evidence from a WhatsApp Deactivation Experiment in Brazil." Working paper.

Wager, Stefan and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113(523):1228–1242.

Weghorst, Keith. 2022. *Activist Origins of Political Ambition: Opposition Candidacy in Africa's Electoral Authoritarian Regimes*. Cambridge University Press.

Zhuravskaya, Ekaterina, Maria Petrova and Ruben Enikolopov. 2020. "Political effects of the internet and social media." *Annual Review of Economics* 12(1):415–438.

A Appendix

A.1 Formalizing effects of access to social media

Here, we propose a simple model to analyze the effects of providing access to social media and the competing effects of imposing a social media ban that can only be circumvented by using a VPN.

A.1.1 Effects of access to social media

We start with the following model of individual i's support for the incumbent over opposition parties:

$$\hat{\mu}_i := \mu_i + \alpha_i(m - \mu_i) + S_i\beta_i(s - \mu_i),$$

where μ_i is i's prior level of support for the incumbent, m is the average support-relevant persuasive signal provided by traditional media, s is the average support-relevant persuasive signal on social media, $S_i \in \{0,1\}$ is an indicator equal to one for individuals who consume social media,[1] and $\alpha_i \in [0,1)$ and $\beta_i \in [0,1)$ are weights capturing the degree to which an individual updates their support from each signal relative to their prior support (whether due to signal credibility or the number of signals received). This reduced-form belief updating model captures a broad set of possible belief models of belief formation, including Bayesian learning with normally-distributed priors (e.g. Arias et al. 2022).

The effect of access to social media for an individual in isolation—in the sense that social media content is not altered by granting access and recipients do not attribute their access to the incumbent—is then given by the difference in support between an individual with and without access to social media:

$$\begin{aligned} \tau_i &:= \hat{\mu}_i(S_i = 1) - \hat{\mu}_i(S_i = 0) \\ &= \beta_i(s - \mu_i), \end{aligned}$$

which is positive (negative) when $s > (<) \mu_i$. As argued in the main paper, $s < \mu_i$ is likely for most individuals in electoral autocracies due to biased or selective coverage. Integrating over the distribution of μ_i (and β_i), we anticipate that the average treatment effect of access to social media, $\mathbb{E}[\tau_i]$, will be negative; empirically, this is identified by randomizing an encouragement for S_i. Moreover, for any given s, the effect of social media is decreasing in prior support μ_i for the

[1] An intensive margin interpretation of a social media treatment that increases the relative weight β_i on social media content yields analogous results.

incumbent. These predictions generate our hypotheses regarding the average and heterogeneous effects of subsidizing access to social media.

A.1.2 Effects of introducing a social media ban

We next turn to the differential effect of imposing a social media ban across VPN and non-VPN users. We assume VPN users maintain access to social media while non-VPN users can no longer access it, although the less stark reality is that non-VPN users face a higher cost of circumventing a ban. Starting from prior support μ_i, expected support for the incumbent among individuals of type $i \in \{V,N\}$ (VPN or non-VPN user) is now given by:

$$\mathbb{E}[\tilde{\mu}_i] := \mu_i + \alpha_i(m - \mu_i) + \mathbb{1}[i = V]\beta_i(s + \Delta_s - \mu_i) - (c + \mathbb{1}[i = N]\Delta_c)$$

where Δ_s captures any difference in social media persuasion due to the ban, c is a common cost of the ban (such as revealing a bad incumbent type to all individuals), Δ_c is the differential cost for individuals who lose access to social media (e.g. upset at lost access, more aware of the ban), and $\mathbb{1}[\cdot]$ is the indicator function. Lost support due to the social media ban is considered as a separate source of support from belief updating about traditional and social media content.

To mimic our empirical analysis of Uganda's social media ban, we take a difference-in-differences approach to examine *relative* changes in support for the incumbent:

$$\begin{aligned}
\mathbb{E}[\tau_{DiD}] &:= (\mathbb{E}[\tilde{\mu}_V] - \mathbb{E}[\mu_V]) - (\mathbb{E}[\tilde{\mu}_N] - \mathbb{E}[\mu_N]) && (4)\\
&= [(\alpha_V - \alpha_N)m + (\alpha_V\mu_V - \alpha_N\mu_N)] + \beta_V(s + \Delta_s - \mu_V) + \Delta_c. && (5)
\end{aligned}$$

The first term captures differences in the degree of updating from other media content, reflecting both differences in the weight attached to the signal reported by traditional media outlets (e.g. $\alpha_V < \alpha_N$ because VPN users are more skeptical of traditional media) and differences in the impact of the signal (e.g. $\mu_N > \mu_V$ because non-VPNs users are already more favorable toward the incumbent, and thus update less from traditional media). However, α_i and $|\alpha_V - \alpha_N|$ are likely to be relatively small due to extensive prior exposure to ruling party messaging and its lack of credibility. The second term captures belief updating from general social media content s together with the change in media content due to the ban Δ_s (e.g. a change in what is available on social media), relative to a citizen's prior belief. The third term captures the differential cost of the ban on non-VPN users. The common component of citizen sanctioning for the social media ban c is not captured by τ_{DiD} because it is absorbed as a common period shock.

Focusing on the latter two terms, the preceding model implies that the difference in the effect of the social media ban on VPN relative to non-VPNs users is theoretically ambiguous. On one hand, as with subsidizing access to social media in a partial equilibrium analysis, typical content on social media is likely to *reduce* support for the incumbent among VPN users relative to non-VPN users to the extent that $s < \mu_V$. This condition may be somewhat less likely to hold than in the general population because VPN users are less favorable toward the incumbent at baseline, but is nevertheless plausible in electoral autocracies where biased traditional media outlets in part shape citizens' prior beliefs.

On the other hand, two new forces (relative to the previous model) counteract the effect of typical social media content to *increase* support for the incumbent among VPN users relative to non-VPN users. First, if $\Delta_s > 0$, then social media content becomes more favorable toward the incumbent during the ban, making it more likely to be the case that $s + \Delta_s > \mu_V$. Second, if non-VPN users face larger costs of the social media ban, then $\Delta_c > 0$ could lead support for the incumbent to drop more among non-VPN users than VPN users. If either effect is sufficiently large, $\tau_{DiD} > 0$, as we ultimately observe empirically.

A.2 Ugandan Facebook data

A.2.1 Corpora of Facebook posts

Our data on Ugandan Facebook posts comes from Crowdtangle. Crowdtangle tracks all Facebook pages with more than 25,000 followers or public accounts (whether *pages* or *groups*) otherwise specifically added to Crowdtangle by researchers. Using this, we construct two corpora of Facebook posts for the 18 month period between June 1, 2020 and December 31, 2021. For each post, we observe covariates including its contents, URL, number of comments, and total number of user interactions. We exclude all posts lacking any text.

For Corpus 1, we extract data relating to a curated set of political *pages* and *groups* with more than 1,000 followers or members, after adding the relevant accounts to Crowdtangle where necessary. Within this corpus, there are multiple account types. First, we extract data from the 145 MP candidates who used a *page* (rather than posting on their private account, for which we cannot extract data).[2] Second, we extract data from a set of 31 prominent Ugandan media outlets. We then code the partisanship of these posts using our contextual knowledge and research assistants. Third, we extract data from a set of *pages* and *groups* about Ugandan politics. These derive from targeted searches of a set of terms within these pages and groups associated with the election—including the

[2] For candidate pages, we looked for the Facebook profiles of up to three most competitive candidates for Parliament in 498 constituencies, yielding 1,430 candidates (some were uncontested, while in some constituencies only two candidates ran). We then found Facebook profiles for 1,363 candidates, 874 (61% of the sample frame) of them with high confidence about the match (likely or certain). Of those, only 282 could be uploaded to Crowdtangle, most likely because the other accounts were private Facebook accounts. From these 282, only 145 accounts posted during the time period of interest.

various party names, political leaders, and government agencies. This generated 221 *pages* and 156 *groups*; the partisanship of these groups was classified by research assistants. Last, we collect data from a set of 31 Facebook pages administered by the Government of Uganda relating to different ministries.

Table A1 breaks down the distribution of these different data sources (columns) and their political affiliations (rows). In total, this corpus provides data on 2.6 million posts throughout the study period. Table A2 documents the total number of posts made on the different types of pages and groups.

Table A1: Distribution of accounts in Corpus 1

Partisanship	MP Candidates	Media	Group	Page
NRM	60	11	50	46
NUP	25	4	93	128
Other opposition	22	12	13	15
Independent	38	0	0	0
Government	0	0	0	32

Table A2: Distribution of posts in Corpus 1

Partisanship	MP Candidates	Media	Group	Page
NRM	4285	206445	323418	13698
NUP	4875	5874	1441254	51846
Other opposition	3142	134042	305553	5875
Independent	1686	0	0	0
Government	0	0	0	10598

For Corpus 2, we extract data from the universe of *pages* tracked by Crowdtangle where the administrator of the page has set Uganda as their country. This comprises a much broader set of relatively popular Facebook pages with, in general, a much lower intensity of political content. We exclude from this corpus the small number of posts also included in our selective Corpus 1.

In total, this corpus provides data on 4.4 million posts throughout the study period from 12,521 distinct Facebook pages.

A.2.2 Classifying Facebook posts

With this aggregate corpus of 6.95 million Facebook posts, we executed three classification tasks to define (A) whether a post is *political* or not, (B) whether a post is *anti-government*, *neutral*, or *pro-government*, (C) the substantive topic of the post. Due to the lack of relevant libraries for classifying Uganda's local languages, we classify the language of every post and restrict to the 83%

of posts coded as English. Further, we preprocess the data to remove user mentions, URLs, and truncate the post text to comprise 128 tokens (roughly similar to a Tweet).

For tasks (A) and (B), we classified posts in two steps using a combination of GPT-based labeling and BERT-based extrapolation, due to the high performance and low costs of BERT when applied to reasonably simple classification tasks (Dell 2024). For task (C), we do this entirely using GPT coding due to the relative complexity of the topic classification task and recent availability of OpenAI's cheaper GPT-4o mini model.

For tasks (A) and (B), we used GPT-3.5-turbo to first label a training set of posts. For task (A), this comprised 10,000 posts which we drew from the subset of Corpus 1 pertaining to MP candidates (we anticipated this should contain a higher share of political content than the rest of our data, and so would avoid issues relating to labeling and predicting rare events). For task (B), this comprised labeling 30,000 posts from the subset of posts classified as *political*. This larger training set owed to the relative complexity of the three-value labeling task plus the rarity of pro-government posts in the corpus overall. Consistent with a number of recent contributions, we find the GPT labeling to perform extremely well (Gilardi, Alizadeh and Kubli 2023; Ziems et al. 2023). Fourth, we used BERTweet, a model pretrained on a large dataset of English Tweets, to classify the full corpora of posts outside of the small set labeled by GPT. After pretraining the model, we obtained accuracy metrics of 85% for task A and 80% for task B, where the "ground truth" is defined by GPT's labeling.

Table A3: Share of anti-government posts in Corpus 1

Partisanship	MP Candidates	Media	Group	Page
NRM	0.15	0.29	0.21	0.21
NUP	0.43	0.56	0.41	0.59
Other opposition	0.31	0.31	0.36	0.37
Independent	0.22			
Government				0.09

Table A4: Comparison of classifications across corpora

Corpus	Political	Anti-government	Neutral	Pro-government
1	0.57	0.36	0.59	0.04
2	0.20	0.11	0.87	0.02

We validate these classifications in two ways. First, we compare the resulting measures to our own coding of the political affiliation of different Facebook groups and pages in Corpus 1 (as in Table A2). In Table A3 we document the share of the different posts classified as being *anti-government*, which strongly correlates with our own coding of these pages and groups. Table A4 intuitively documents a much higher share of *political* posts in Corpus 1 (57%) than in Corpus 2 (24%).

Table A5: Validating classification of Facebook posts

		All		Corpus 1		Corpus 2	
		Accuracy	F1	Accuracy	F1	Accuracy	F1
		(1)	(2)	(3)	(4)	(5)	(6)
A. Classifying posts as political							
GPT+BERT	Coder 1	0.80	0.85	0.74	0.77	0.85	0.91
GPT+BERT	Coder 2	0.83	0.88	0.81	0.82	0.85	0.91
Coder 1	Coder 2	0.91	0.94	0.87	0.90	0.94	0.97
B. Classifying posts as anti-government							
GPT+BERT	Coder 1	0.65	0.69	0.69	0.68	0.60	0.69
GPT+BERT	Coder 2	0.72	0.70	0.73	0.67	0.69	0.73
Coder 1	Coder 2	0.82	0.85	0.82	0.82	0.83	0.88

Second, two Ugandan research assistants hand-coded identical sets of 1,000 posts, randomly drawn from the two corpora, to replicate the two classification tasks. Table A5 provides measures of coding similarity between our GPT+BERT measure and the two coders, as well as between the two coders themselves. We provide measures of accuracy and F1 for the combined corpus and each corpus individually. Especially given the complexity of the classification tasks, we find our classifications to perform well. In panel A, we find accuracy measures of between 0.80 and 0.83 between our GPT+BERT measure and the two coders when classifying posts as containing political content. This accuracy measure is slightly higher in Corpus 2, which contains a lower share of political content to begin with. In panel B, we find slightly lower accuracy scores on the anti-government sentiment classification task of between 0.65 and 0.72, with greater accuracy in the more-political Corpus 1. Patterns using the F1 score are similar. Importantly, these figures should be benchmarked against the extent of agreement between our coders, which for the accuracy measure was 0.91 for political classification and 0.82 for the more complex task of classifying sentiment. These benchmarks suggest that our GPT+BERT classification was only around 10-15% less accurate than a hand-coded benchmark while covering a vastly greater scale of data.

For task (C), we used the recently-released Open AI's GPT-4o-mini to classify the topic of posts. The greater complexity of this topic classification task, plus the relatively low cost of GPT-4o mini relative to its performance, meant that we used GPT-4o-mini to classify every post in our corpus rather than training a BERT model. Our basic topics were initially constructed by reviewing citizens' perceptions of the most important issues facing Uganda in Afrobarometer (2019), before supplementing these with apolitical topics relating to entertainment, culture, and personal news which feature prominently in our corpus. For every post, we therefore asked GPT to classify which of the following topics was most relevant, where a given post could be assigned at most to three topics:

1. political parties, politicians, candidates, campaigns;
2. voting, elections, rallies;
3. election fraud, rigging, vote buying;
4. government, government ministries, government policies;
5. crime, security, justice, prisons, police, military;
6. corruption, bribery, embezzlement;
7. protest, unrest, riot;
8. human rights, journalism, freedom of speech, civil liberty;
9. land, expropriation, property rights;
10. economy, wages, labor, inflation, currency, poverty, unemployment;
11. farming, agriculture, crops;
12. science, technology;
13. weather, environment, extreme weather, natural disasters, climate change;
14. social programs, policies to reduce poverty;
15. education, exams, schools, universities;
16. health, public health, healthy living, health care;
17. infrastructure, roads, transport, cars, traffic, traffic accidents;
18. water, sanitation; electricity, power;
19. arts, culture, society, music, festivals, food, fashion, lifestyle;
20. entertainment, celebrities, showbiz;
21. sports, athletics;
22. careers, individual business activities;
23. personal stories, life updates;
24. religion, church, worship, gospel;
25. advertisements;
26. news unrelated to Uganda;
99. other or cannot be classified.

With these codings, we then define the following topical groupings for parsimony:

1. *Election*: 1, 7, 8;
2. *Regime*: 2, 4, 5, 6, 18;
3. *Economy*: 3, 14, 16, 17, 19;
4. *Services*: 9, 10, 11, 12, 13, 15;
5. *Arts and entertainment*: 20, 21, 23;
6. *Personal and religion*: 22; 25; 26;
7. *International news*: 27;
8. *Other*: 24, 99.

While we do not validate these classifications in the same way as for tasks (A) and (B), given their purely descriptive use in the paper, we do find that the classifications vary in intuitive ways with our independent coding of political content and sentiment. In Table A6 we show that nearly all *Election* and *Regime* posts are coded as being political, and most are coded as anti-government; further, these levels are higher in the more-political Corpus 1. Figure A7 further disaggregates the individual topics comprising the *Election* and *Regime* groups.

Table A6: Validating classification of topics

	All		Corpus 1		Corpus 2	
	Political	Anti-govt	Political	Anti-govt	Political	Anti-govt
	(1)	(2)	(3)	(4)	(5)	(6)
Election	0.93	0.67	0.96	0.70	0.87	0.58
Regime	0.85	0.68	0.91	0.78	0.75	0.55
Economy	0.50	0.19	0.73	0.29	0.39	0.14
Services	0.44	0.14	0.62	0.21	0.36	0.11
Arts and entertainment	0.15	0.06	0.35	0.15	0.10	0.04
Personal and religion	0.12	0.04	0.27	0.10	0.08	0.03
International news	0.51	0.10	0.75	0.25	0.44	0.06
Other	0.07	0.12	0.15	0.16	0.04	0.10

A.3 Constructing our panel survey sampling frame

The districts we sampled comprise Mpigi, Kalungu, Masaka in Central region; Iganga, Jinja, Mbale, Sironko in Eastern region; Gulu, Lamwo, Lira in Northern region; and Rukungiri in Western region. Peri-urban trading centers were selected with good 3G internet reception, which we assessed using administrative locality data and the Collins World Explorer database.

In these trading centers, we designed a tiered phone-based recruitment process to construct the sampling frame. First, we obtained contact details of community leaders—including LC1 councilors, parish chiefs, and village health team members—within a given TC from district-level officials. Second, in phone calls with these leaders, we obtained a list of up to eight "seeds" per trading center (TC) stratified by their role (*boda boda* drivers, teachers, business-persons, or youth representatives) within the community. Third, we called every "seed" to solicit contact details for a set of their personal contacts who might be interested in, and eligible for, the study. This process generated a sampling frame of 4,399 contact phone numbers for potential respondents for the study across 135 TCs.

A.4 Field experiment design

A.4.1 Randomization and treatment conditions

To randomize treatment and control conditions across respondents first enumerated in the baseline survey ($n = 1,310$), we first created 22 district $\times$ cell phone network strata. Within each strata, we then created nested blocks of eight, four, and two individuals based on a vector of predetermined covariates—including measures of their social media usage, COVID-19 knowledge, the extent of their social interactions online and offline, subjective welfare, and attitudes towards the ruling NRM party—and assigned treatment within the matched pairs. The nested blocks were chosen to allow for the inclusion of block fixed effects in the presence of attrition that removed within-block variation in treatment. For the residual sample first recruited for the midline survey ($n = 145$), for whom we lacked covariates observed on the baseline survey, we assigned treatment within strata using complete randomization within the midline survey. In each case, treatment—the form of compensation for completing the survey—was assigned during the midline survey.

Our treatment condition varied slightly over the study period due to the elimination of the OTT tax. In June, we: (i) paid the monthly OTT tax (UGX 6,000); and (ii) provided 1.5GB of mobile data for the month for Airtel users (UGX 10,000) or 500MB a week for MTN users (UGX 5,000 a week for four weeks). In July and August, treated Airtel users received 400MB a week (UGX 3,500 a week) and treated MTN users received 500MB a week (still costing UGX 5,000 a week). SMS messages informed participants of each transfer. Individuals assigned to the control condition were instead compensated with a flexible mobile money transfer of UGX 6,000 shortly after the wave 2 survey enumeration.

A.4.2 Automated selection of covariates

For the LASSO-selected covariates, we considered the superset of all potential covariates, $\mathbf{X}_i^+$, from the wave 1 or 2 surveys with full data coverage along with trading center fixed effects (which we prespecified as an auxiliary specification). Following Belloni, Chernozhukov and Hansen (2014), $\mathbf{X}_i$ is the union of covariates selected by LASSO when $Treatment_i$ is predicted by $\mathbf{X}_i^+$ and Y_i^{post} is predicted by $\mathbf{X}_i^+$. In some cases, this procedure did not result in the selection of additional covariates.

A.4.3 Deviations from pre-analysis plan

We registered the experiment and our pre-analysis plan at the AEA registry, which is available here. The estimation of average and heterogeneous treatment effects and variable construction adhered to prespecified procedures, with the following exceptions:

1. Whereas the pre-analysis plan specified the use of one-tailed tests for directional hypothesis, we use two-tailed tests throughout our analysis. This more conservative inferential strategy matches the observational analysis.

2. We prespecified using the randomization blocks of size 4 for our fixed effects, but ultimately decided to use blocks of size 8. While this choice has little consequence for the estimation of average treatment effects, we did so to minimize the number of block $\times$ NRM supporter

groups containing no variation in treatment (due to attrition and limited baseline variation in prior NRM support). Table A16 demonstrates that our results are robust to including fixed effects for blocks of size 2, 4, and 8 as well as including no block fixed effects at all. Our point estimates increase as more fine-grained blocks are used, but this also entails a loss of precision as more observations from blocks lacking variation in treatment are effectively dropped.

3. We prespecified that respondents who refused to respond to questions would be dropped from the analysis. Because these were very rare and not consistent within respondents across the questions comprising our key outcome indexes, we instead followed our prespecified protocol for imputing don't know responses. Again, this minor deviation is of little consequence: Table A15 reports very similar results when we drop the handful of respondents who refused to answer outcome questions from the analysis.

4. We slightly adjusted the computation of two political outcome measures, but without affecting the substance or significance of our estimates. With respect to political support, we prespecified that our outcome index would use differences in NRM support relative to opposition parties for the thermometer and vote openness outcomes. To provide more insight about how social media access altered political support across parties, we instead created separate indexes focusing on NRM and opposition support that used the thermometer and openness outcome levels for each party instead of the differential between government and opposition party. Unsurprisingly, given that NRM support decreased and opposition support increased among prior NRM supporters, Figure A8c and panel C of Table A10 reports similar results using the prespecified index. With respect to government performance approval, we prespecified using an index containing five items. But our analyses in the main paper only focus on the central government performance item in order to match our observational analyses, since the other four items were not measured in the wave 1 and 2 surveys used in the observational study. Again, Figure A8b reports similar experimental results when using the full index, with drops in performance approval among NRM supporters reflecting lower perceptions of central government performance, trust in government information, and agreement with the social media tax.

5. We prespecified that we would examine self-reported measures of social media use as well as our behavioral observation of WhatsApp use. We ultimately excluded the self-reported survey responses from our analysis because respondents struggled to accurately report the number of hours that they used different media platforms in an average week over the three months prior to the endline survey. This issue, of course, does not affect our behavioral measure of usage. Other studies encounter similar challenges with noisy self-reported social

media use data compared to behavioral measures (e.g. Guess et al. 2023). We report results using these noisy self-reported measures in Table A17, which provides evidence of positive treatment effects on social media usage, on both extensive and intensive margins, for all of Facebook, WhatsApp, Facebook Messenger, and Twitter.

6. We prespecified that the prior NRM supporter moderator would be defined, based on wave 2 survey responses to local election choices, by voting for the NRM candidate for MP (we asked for the candidate name to avoid sensitivity about asking about national vote choices) and for LC5/district chairperson (we asked about party of candidate). Unfortunately, neither measure turned out to be a suitable moderator due to substantial non-response: 55% for MP vote choice and 24% for LC5 chairperson. Non-responses likely reflect respondents' inability to name the candidate for MP they voted for and unwillingness to report on this sensitive issue, which is itself likely to be a function of partisanship; the accuracy of reported vote choice may also be suspect for the same reason. As noted in the main paper, we instead use wave 2 measures to define prior NRM support as an indicator for respondents who reported both feeling more warmly towards the NRM and being more open to voting for the NRM than any opposition party. In addition to substantially reducing non-response, this alternative measure closely mirrors outcomes and is more likely to capture partisan identification with the national NRM party—the primary focus of our study.

We also note that this article restricts attention to political outcomes—knowledge, beliefs, engagement, and ultimately support. A separate article will report results pertaining to the effects of social media access on health outcomes—knowledge and behaviors relating to COVID-19, subjective mental health, and evaluations of social media.

A.5 Research ethics

The design of our study reflected careful attention to the ethics of field experimentation and associated data collection, in line with the American Political Science Association's *Principles and Guidance for Human Subjects Research* (American Political Science Association 2020).

With respect to the data collection, this manifested in three main ways. First, in collaboration with our local enumeration firm and as approved by a local IRB, our instrument was designed to avoid forcing the disclosure of potentially politically sensitive information by respondents. This included not asking respondents directly about their vote choices in the presidential election, which would also not have been permitted by the local IRB, and always allowing them to refuse to answer potentially sensitive questions. The survey instruments themselves, therefore, were designed to minimize any risk to respondents. Second, since enumeration took place during the COVID-

19 pandemic, the decision to remotely enumerate respondents by telephone reflected an effort to minimize any risk to enumerators (though phone-based surveys are quite common in this context to begin with). Third, the behavioral measure of WhatsApp usage was not discussed with respondents. Doing so would have risked introducing Hawthorne effects, especially during the social media ban period when WhatsApp was only accessible by VPN. This data is public-facing by default on WhatsApp and inaccessible for the portion of respondents who had modified the privacy settings of their account. Importantly, we can only observe *whether* a respondent used WhatsApp on a given day, rather than observing anything they did on the platform.

With respect to the experimental intervention, we note that this is a setting where users are constrained in the extent of their social media use by (partially government-imposed) high costs of access. Only 3% of our respondents cited a lack of interest in limiting their greater use of social media platforms; by contrast, 68% reported data costs, 55% the OTT tax, and 18% network coverage. In contrast to Global North settings, where many users may be using social media *above* the socially optimal amount (Allcott et al. 2020), the centrality of online platforms in Uganda for social and economic activities—especially during the COVID-19 pandemic—meant that increased usage could be welfare-enhancing in this setting. We note, however, that our treatment only experimentally facilitated *access* and respondents were free to change their actual *usage* as they liked.

A.6 Reconciling results based on complier characteristics

The different results across our field experimental and observational analyses are unlikely to be driven by compositional differences in *whose* access was shifted by the respective sources of variation in the two research designs.

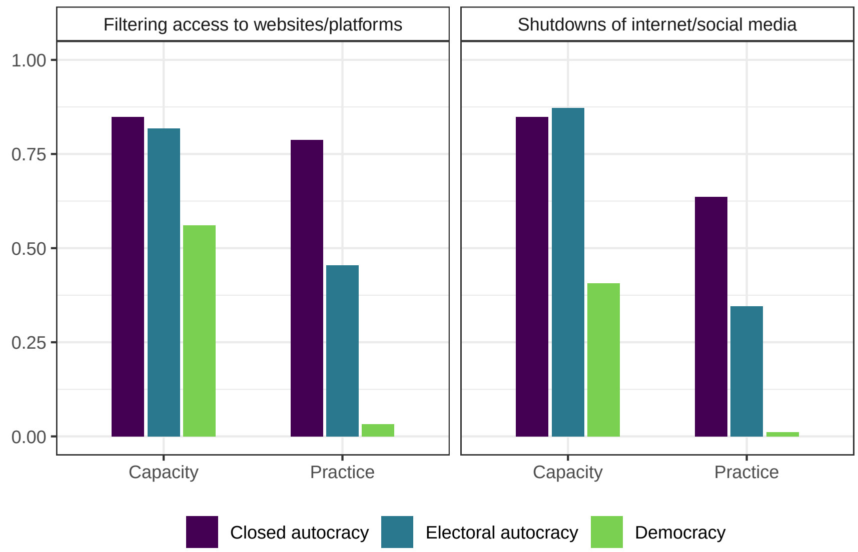
First, we consider mechanical differences in the characteristics of the estimating samples. For one, Table A13 shows that estimating the field experimental results among the subset of respondents who reported using a VPN produces similar—if not, as we note above, slightly stronger—results. For another, while our sample for the field experimental and difference-in-differences analyses do not perfectly overlap, Table A7 demonstrates that they are indistinguishable in terms of their baseline characteristics and attitudes.

Second, we consider subtler differences in the characteristics of the effective "compliers" across the two designs—i.e., whether different types of participant were induced to increase their social media usage across the designs—which might then explain differences in their subsequent political attitudes. Figure A10 compares the characteristics of these effective compliers using the predicted change in participants' WhatsApp usage deriving from the causal forest exercises we describe above. These characteristics broadly overlap, but do show some differences: for example, those induced to increase their usage the most in the field experiment were relatively less educated, while in the difference-in-differences analysis they were relatively more educated.

In the spirit of Angrist and Fernandez-Val (2013), Aronow and Carnegie (2013), and Hotz, Imbens and Mortimer (2005), we assess whether these differences are plausibly sufficient to explain the different results. We do this by weighting the estimating equation associated with each design by the inverse of that respondent's predicted "first stage." Figure A15 presents the results. This exercise, effectively extrapolating the estimated effects beyond the set of compliers whose social media usage was most affected in each design, provides no evidence that the relatively marginal differences in complier characteristics can explain the contrasting treatment effects.

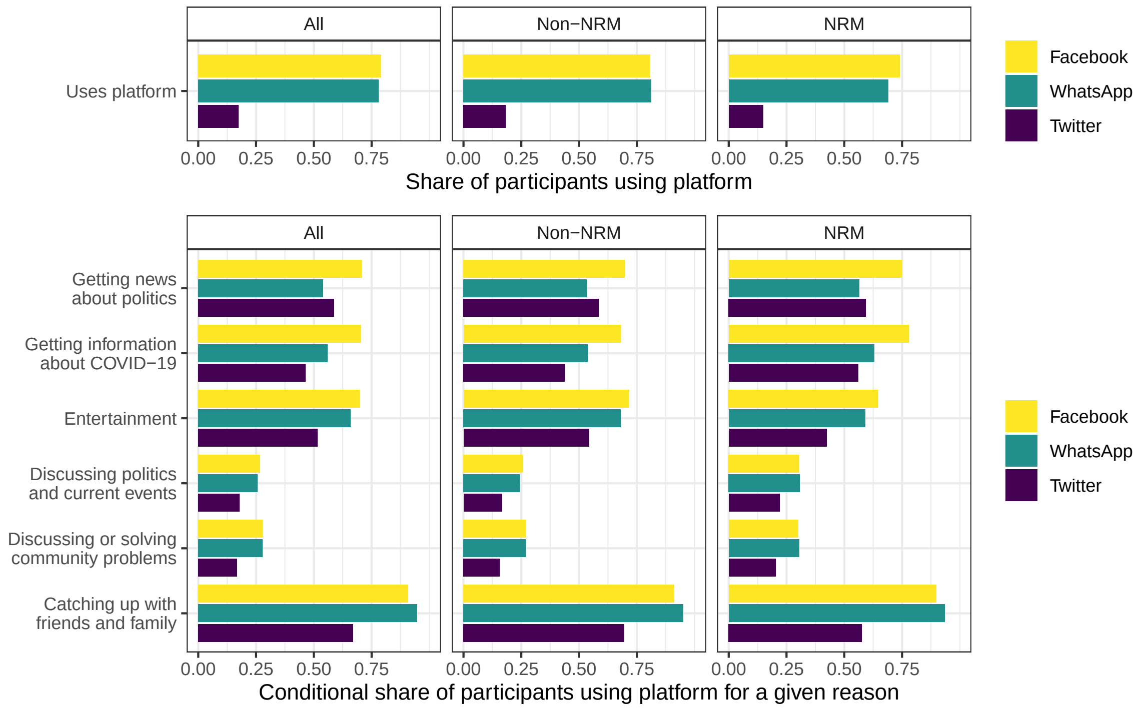
A.7 Additional Figures

Figure A1: Regime types and decisions to restrict online access



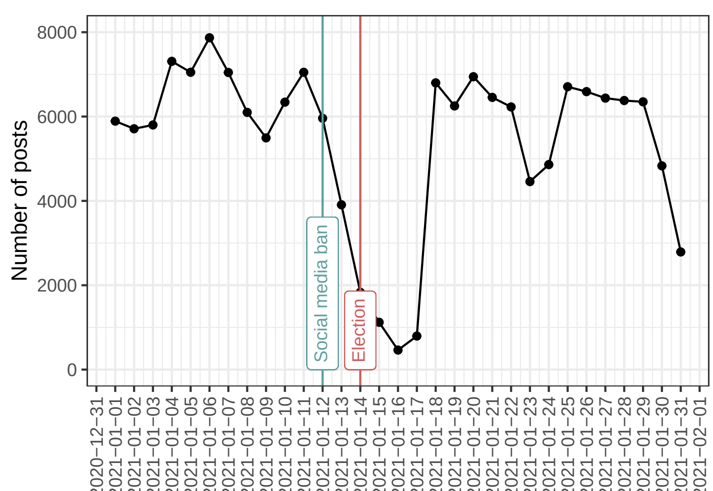
Notes: Figure uses data from 2023 edition of *Digital Society Project* (see Mechkova et al. 2022). V-Dem regime classifications used, aside from pooling *Electoral democracies* and *Liberal democracies*. Filtering capacity: Government has technical capacity to censor information by blocking access to most or all websites/platforms if it decided to do so. Filtering practice: Government does so in practice at least sometimes. Shutdown capacity: Government has technical capacity to shut down most or all domestic access to the internet and social media. Shutdown practice: Government does so at least sometimes.

Figure A2: Usage of social media platforms and reasons for usage in wave 1 survey



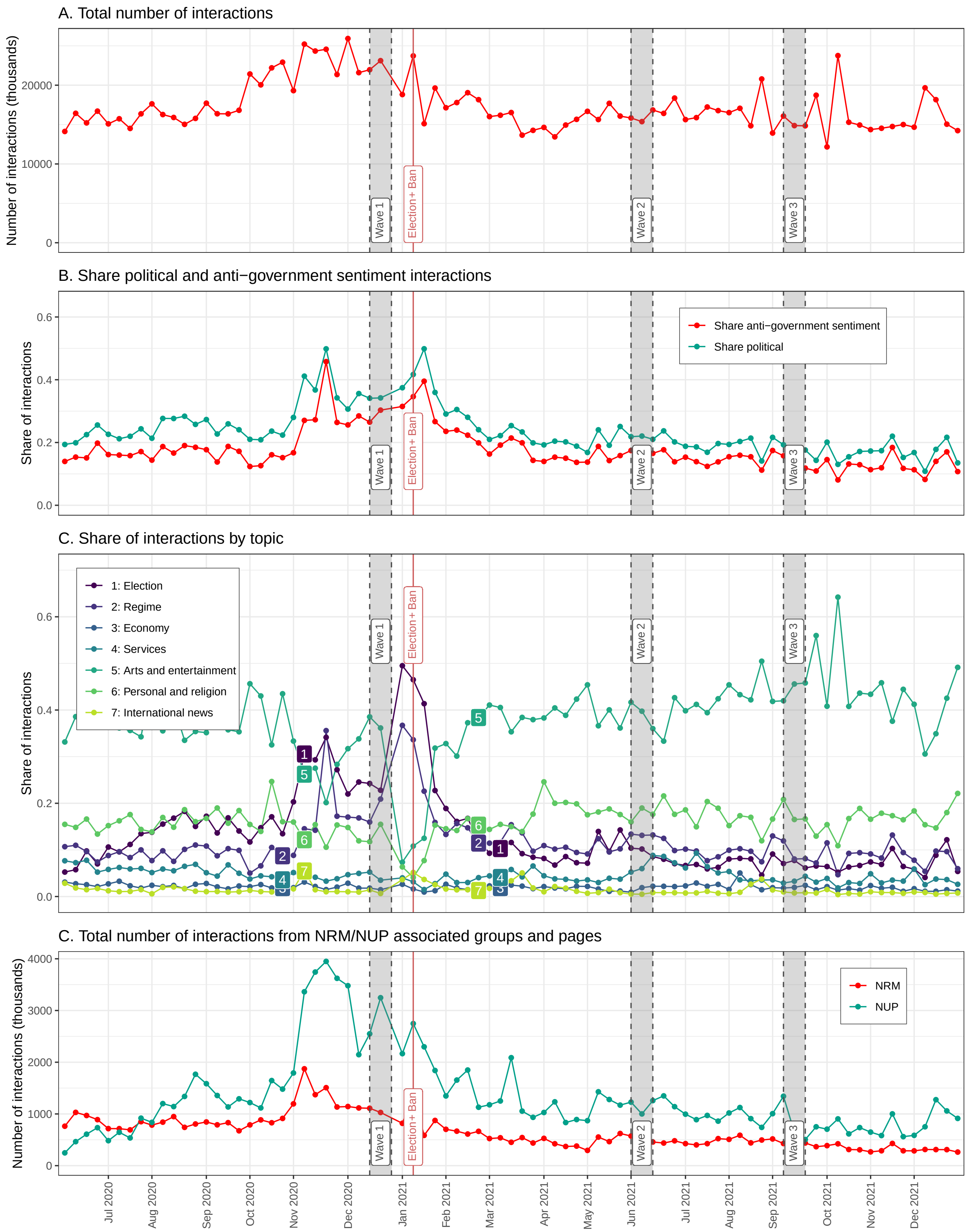
Notes: The top panel reports participants' self-reported use of different social media platforms, overall and by partisanship. The bottom panel reports the main reasons for using a platform conditional on those who report using a given platform.

Figure A3: Day-level Facebook posts around election day



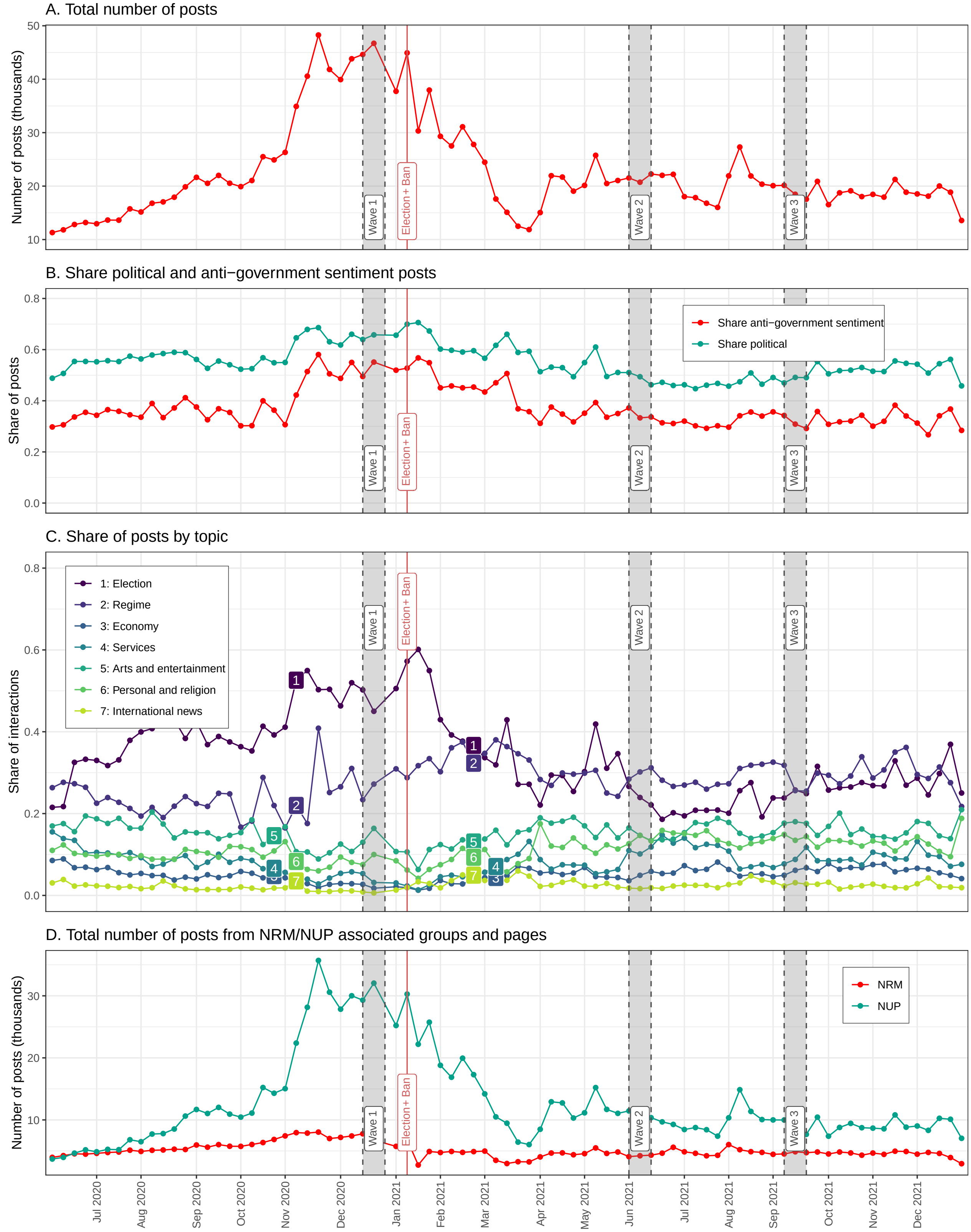
Notes: See Appendix A.2.1 for information on corpus of posts and Appendix A.2.2 for information on coding. The social media ban was introduced on January 12, the internet blackout began on January 13, the election was held on January 14, and the internet blackout was lifted on January 18.

Figure A4: Facebook interactions data during study period



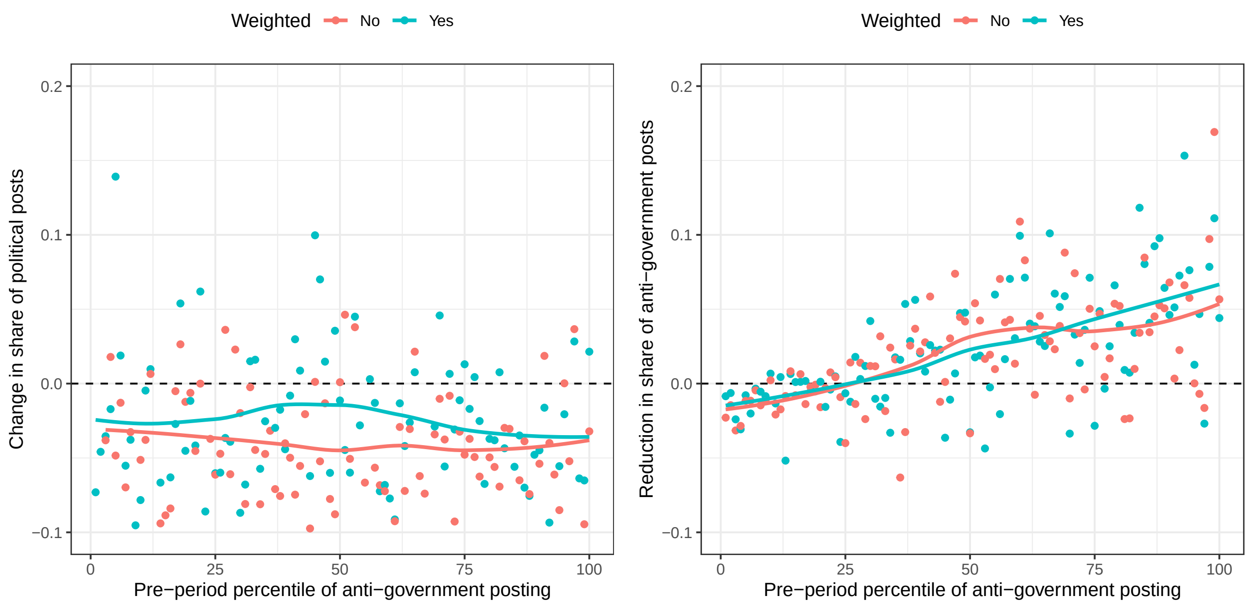
Note: See Appendix A.2.1 for information on corpus of posts and Appendix A.2.2 for information on coding. Crowdtangle-defined 'interactions' comprise likes, reactions, comments and shares.

Figure A5: Facebook post data during study period (corpus 1)



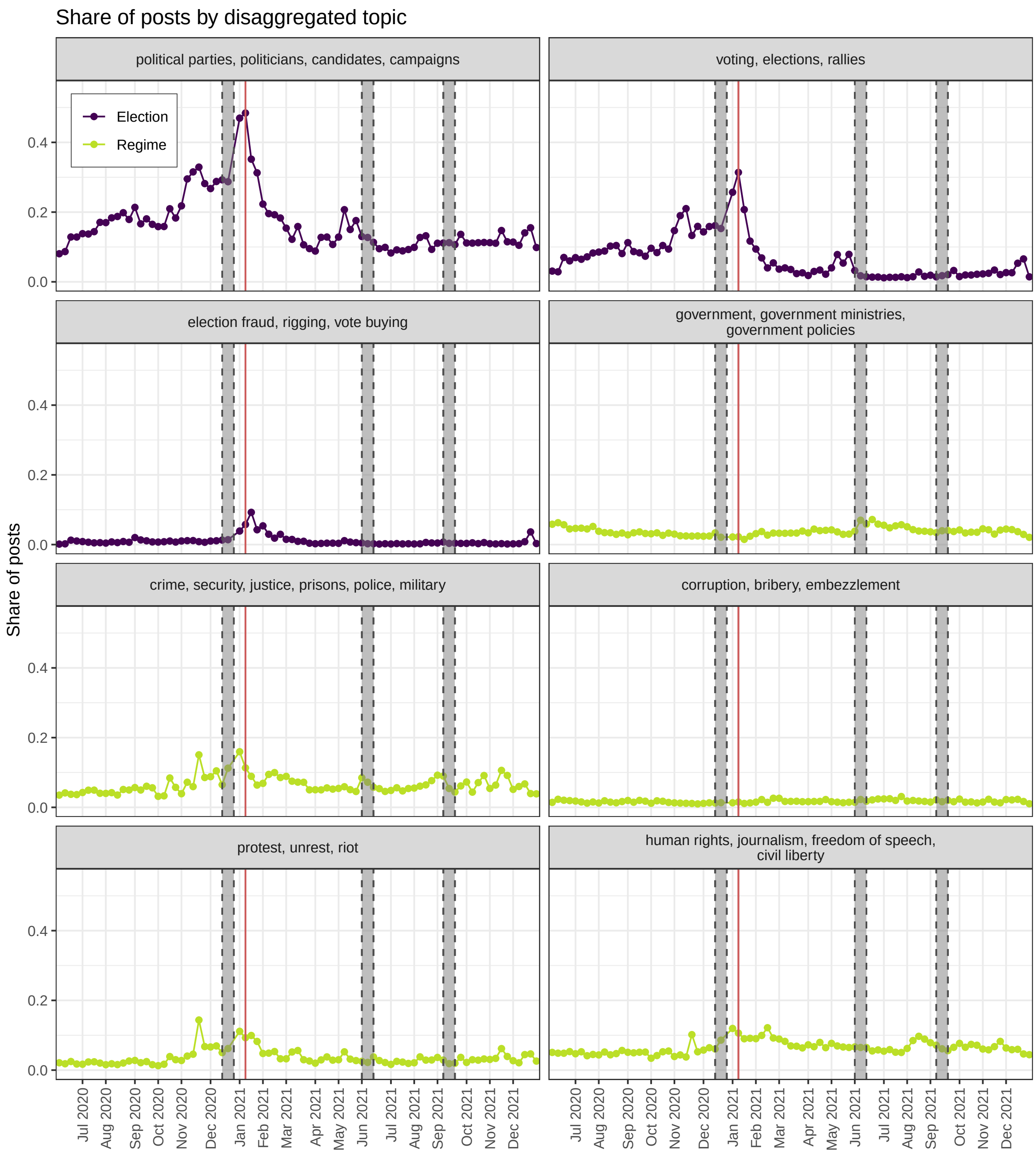
Note: See Appendix A.2.1 for information on corpus of posts and Appendix A.2.2 for information on coding.

Figure A6: Change in posting behavior during ban period based on accounts' pre-ban sentiment

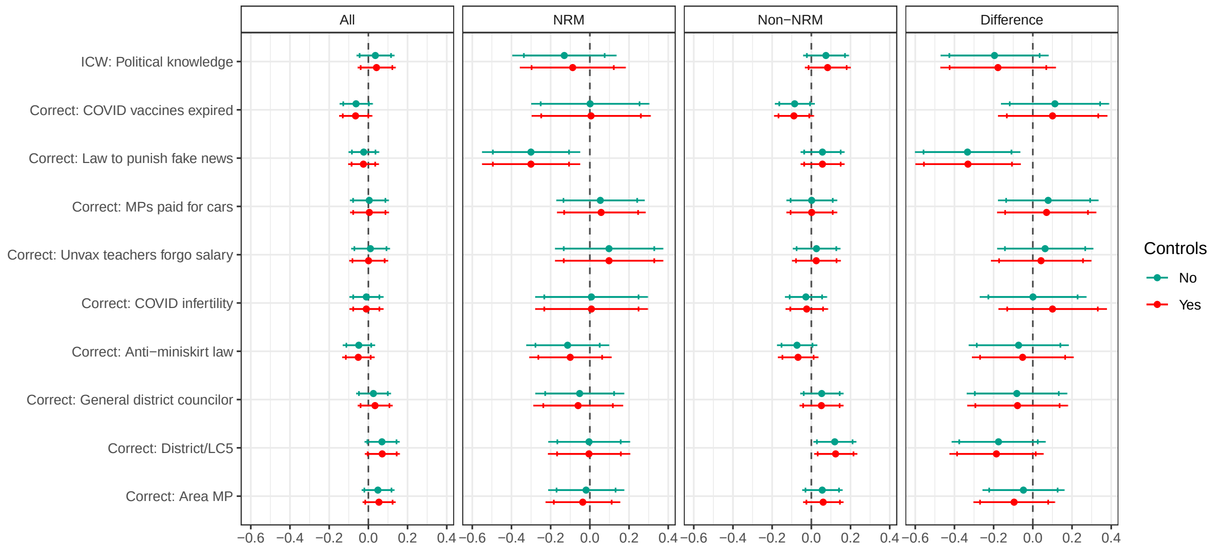


Notes: Figure computed pre-period percentile of anti-government posting by aggregating posts to the Facebook account level in the pre-ban period. Change variables (either reduction in political posts or reduction in anti-government posts) are computed both without weights and using weights based on how many posts a given account made during the ban period.

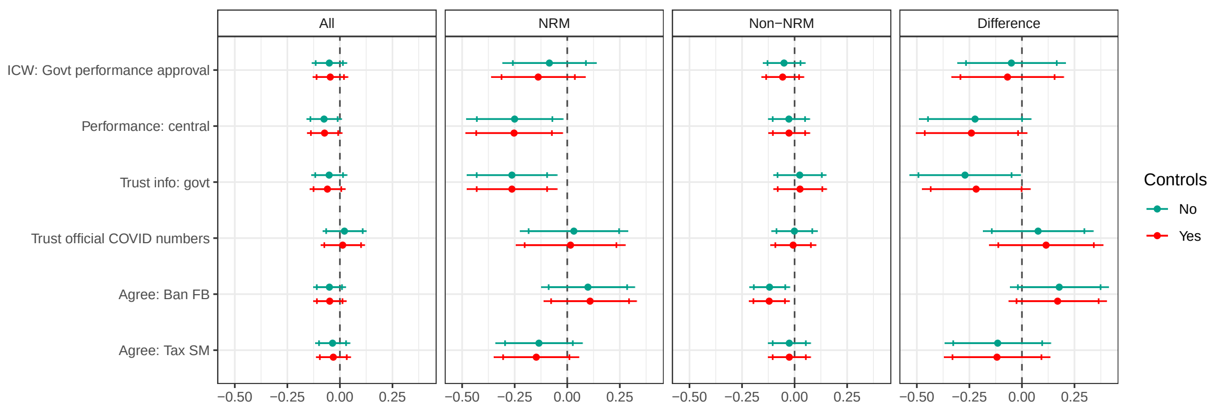
Figure A7: Disaggregation of Facebook post topics



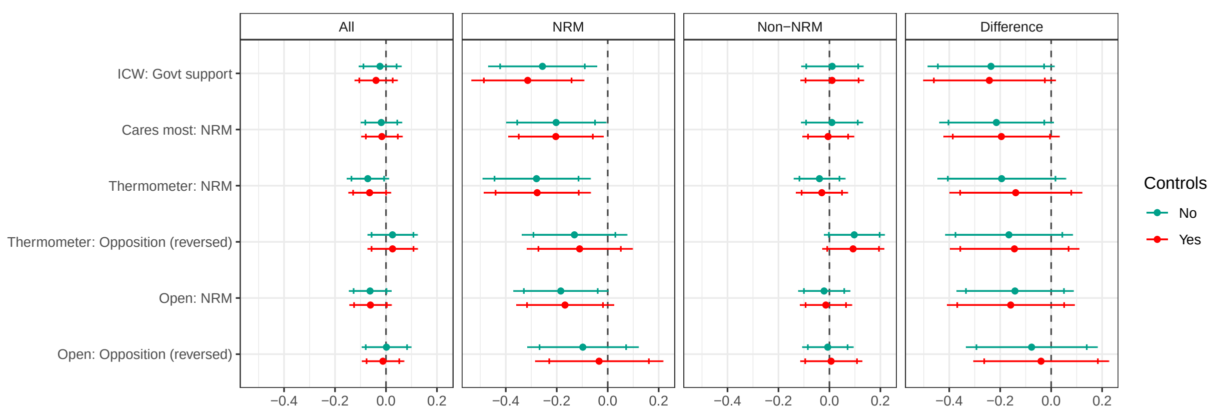
Note: See Appendix A.2.1 for information on corpus of posts and Appendix A.2.2 for information on coding.



(a) Knowledge of current political events



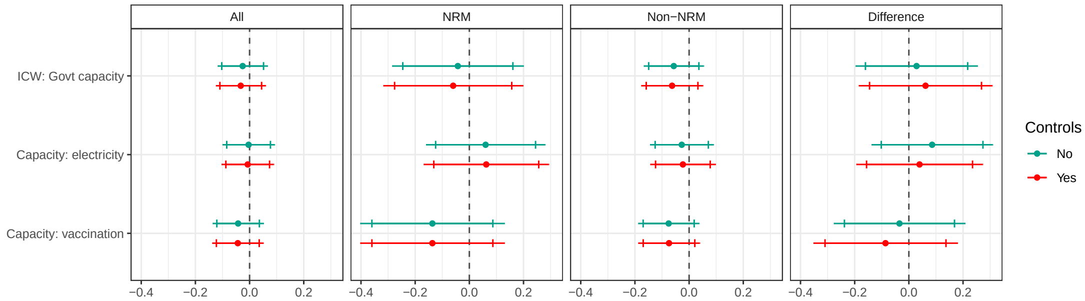
(b) Government performance approval



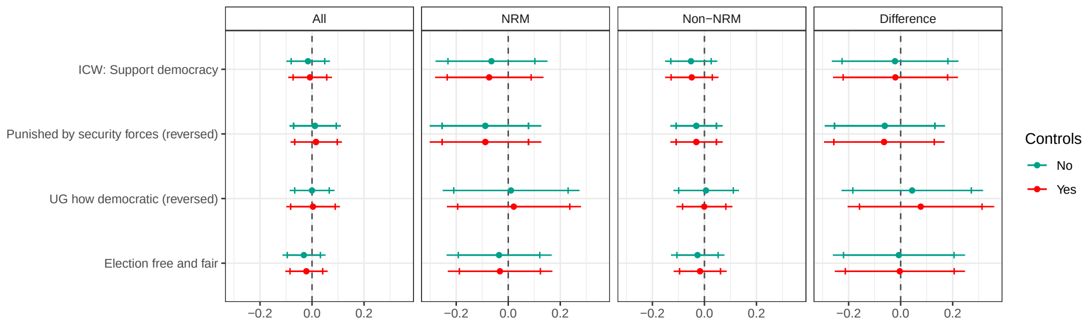
(c) Government support

Figure A8: Treatment effects of experimental social media subsidy on pre-registered indices

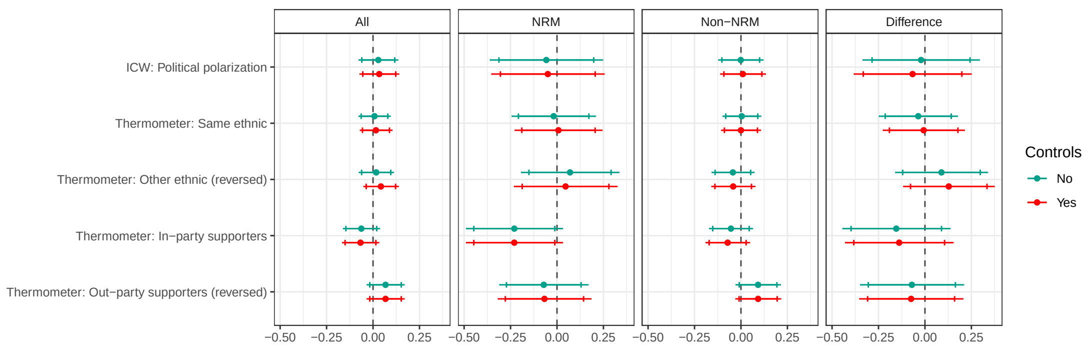
Notes: The estimates in each panel derive from equation (1) estimated in the full sample (left column) and subsamples according to NRM and non-NRM partisanship (middle and right columns). Index outcomes are standardized; subcomponents are unstandardized. The addition of controls indicates adjustment for LASSO-selected covariates. 90% and 95% confidence intervals plotted.



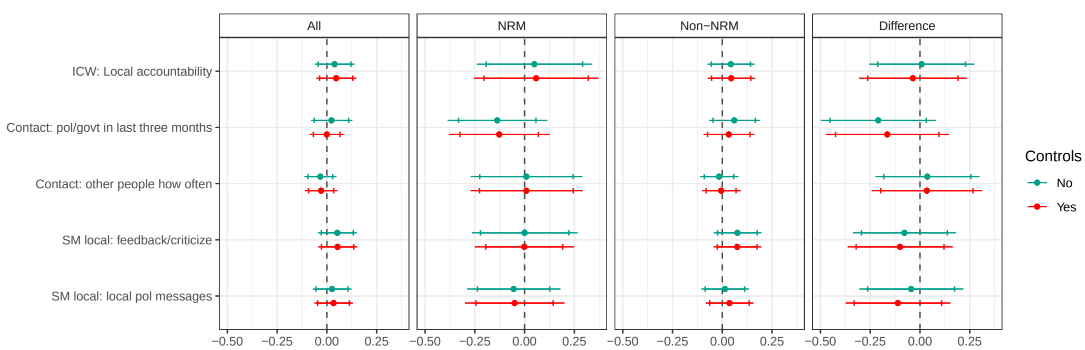
(d) Perceived state capacity



(e) Perceptions of democracy



(f) Affective political polarization

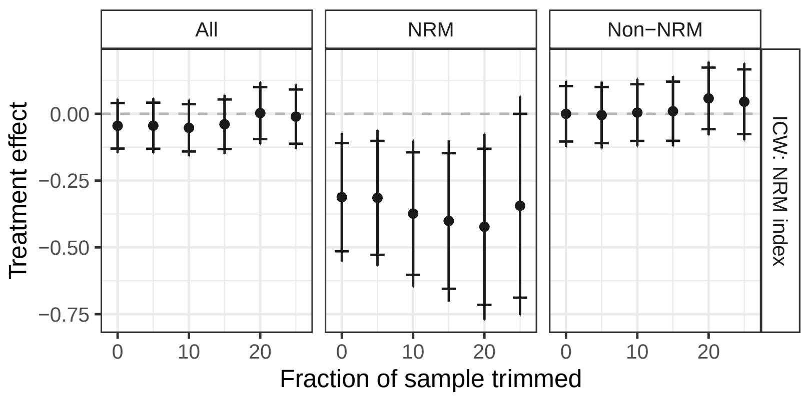


(g) Local accountability efforts

Figure A3 (cont.): Treatment effects of experimental subsidy on pre-registered indices

Notes: The estimates in each panel derive from equation (1) estimated in the full sample (left columns) and subsamples according to NRM and non-NRM partisanship (middle and right columns). Index outcomes are standardized; subcomponents are unstandardized. The addition of controls indicates adjustment for LASSO-selected covariates. 90% and 95% confidence intervals plotted.

Figure A9: Conditional average treatment effect of experimental social media subsidy, by predicted effect on WhatsApp usage



Notes: The estimates in each panel derive from equation (1) estimated among the full sample (first column) and partisan subsamples (second and third columns), where each estimate trims increasing proportions of the sample according to the percentile of their predicted treatment effect on WhatsApp usage (as described in the main text). 90% and 95% confidence intervals plotted.

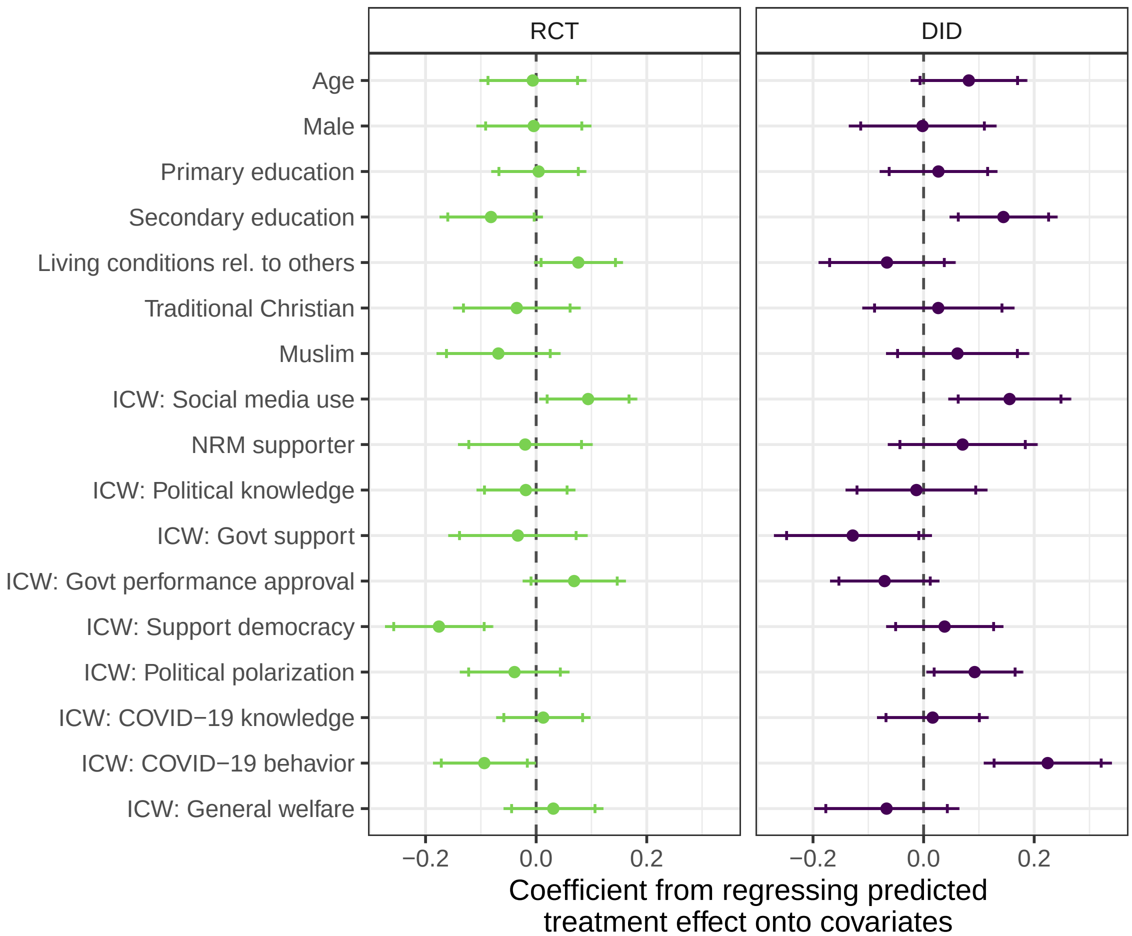
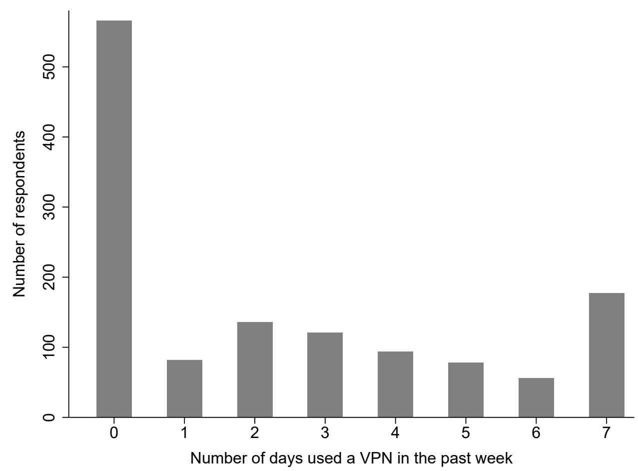


Figure A10: Correlation between baseline survey covariates and predicted WhatsApp usage in the experimental and difference-in-differences research designs

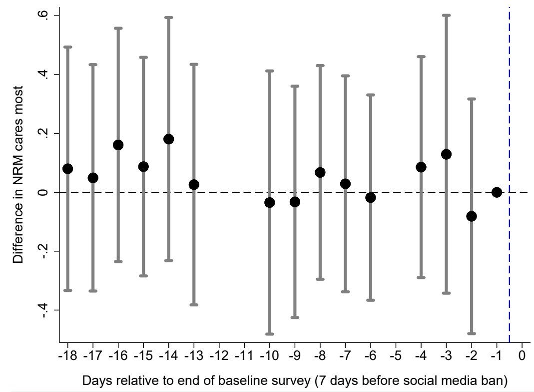
Notes: Both the RCT and difference-in-differences specifications are estimated using OLS including block fixed effects, where standardized baseline covariates differ by row. The RCT specification regresses the predicted treatment effect on WhatsApp usage from field experiment onto the set of standardized wave 1 or 2 covariates (whichever was more recent). The DID specification regresses the predicted treatment effect on WhatsApp usage from difference-in-differences design onto the set of standardized wave 1 covariates. Standard errors clustered by trading center are in parentheses. 90% and 95% confidence intervals plotted.

Figure A11: Distribution of VPN use across wave 1 survey respondents

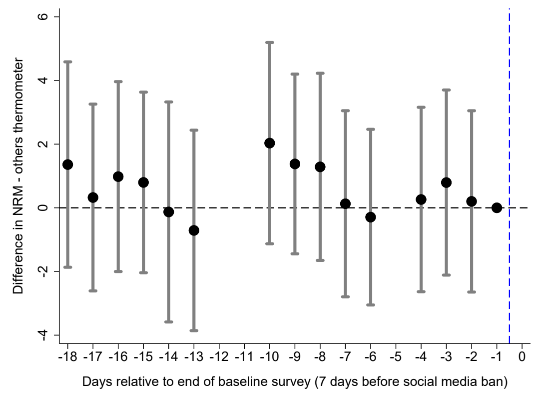


Note: Baseline survey administered in December and January 2020 asked respondents how many days they had used a VPN in the prior week.

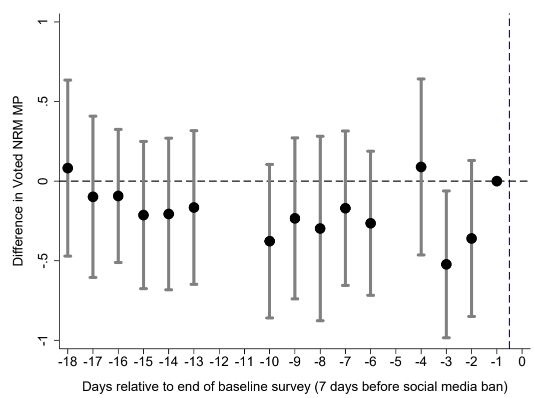
Figure A12: Event study trends in NRM support by baseline survey enumeration date



(a) NRM cares most about people like the respondent

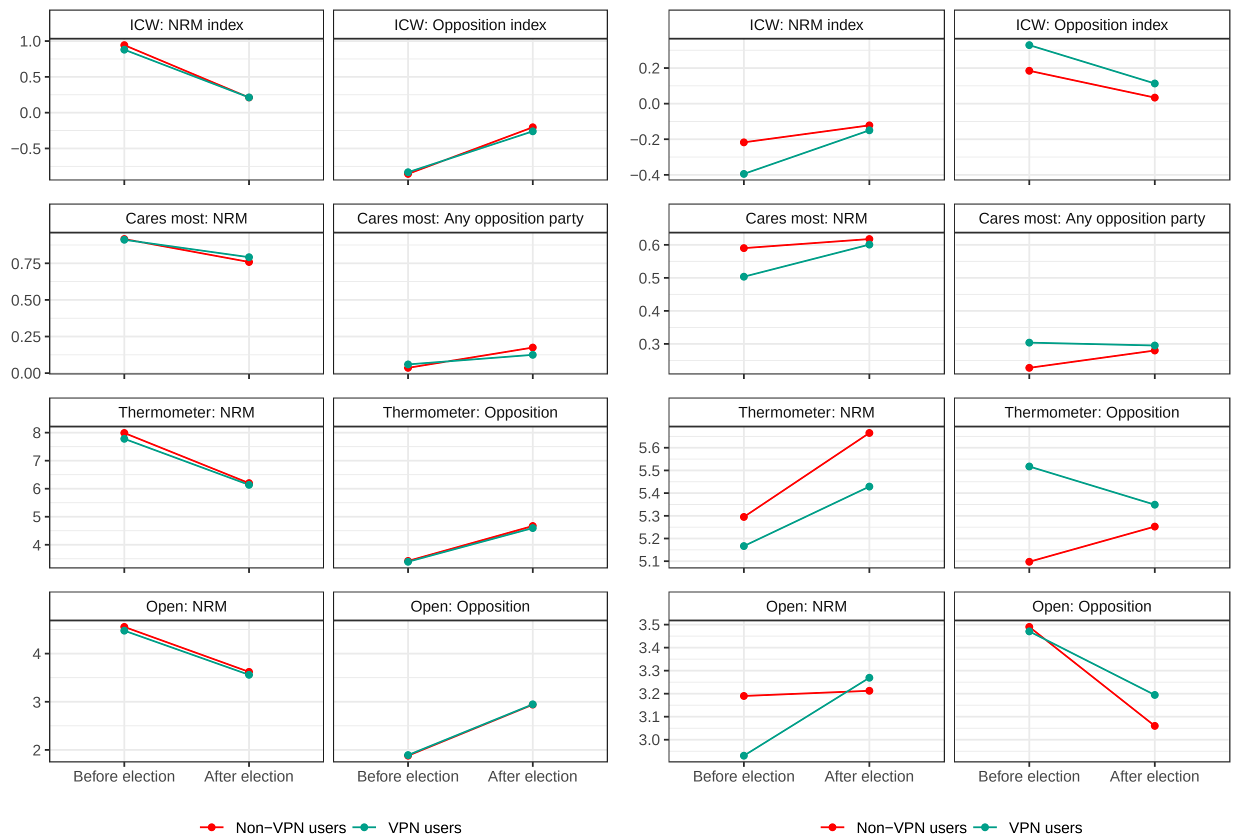


(b) NRM feeling thermometer



(c) Openness to voting NRM

Notes: We include all enumeration days where at least 25 surveys were completed. The reference category is the last day of survey enumeration (seven days before the social media ban).



(a) NRM supporters

(b) Non-NRM supporters

Figure A13: Raw difference-in-difference plots by prior partisanship

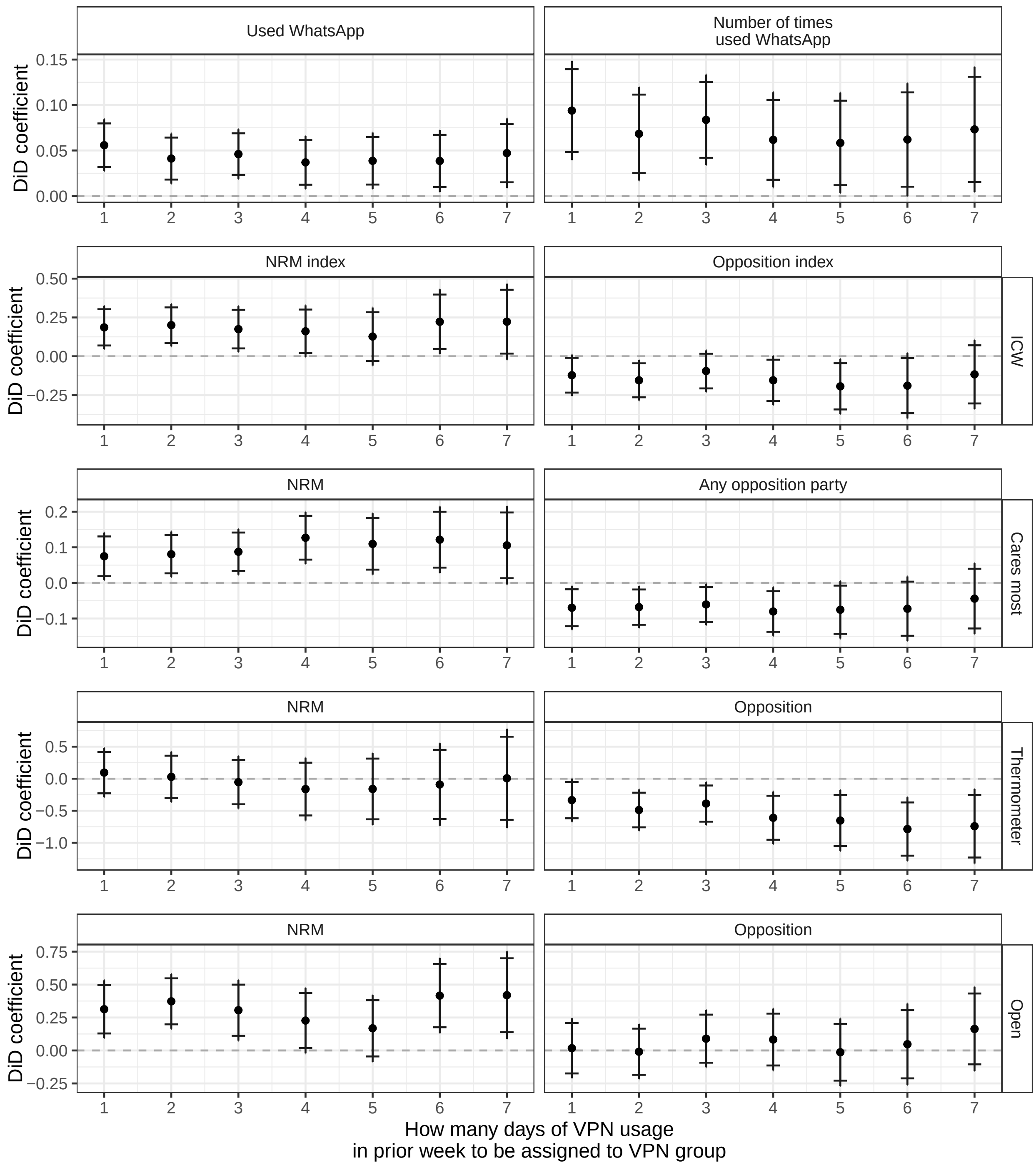


Figure A14: Varying definition of VPN group assignment based on frequency of usage

Notes: Treatment effects estimated using equation (2) including individual and time period fixed effects, with standard errors clustered by district. 90% and 95% confidence intervals plotted. The *x*-axis varies the definition of VPN group based on the number of days in the week prior to the baseline survey that the respondent reported using a VPN; for the distribution of this variable, see Figure A11.

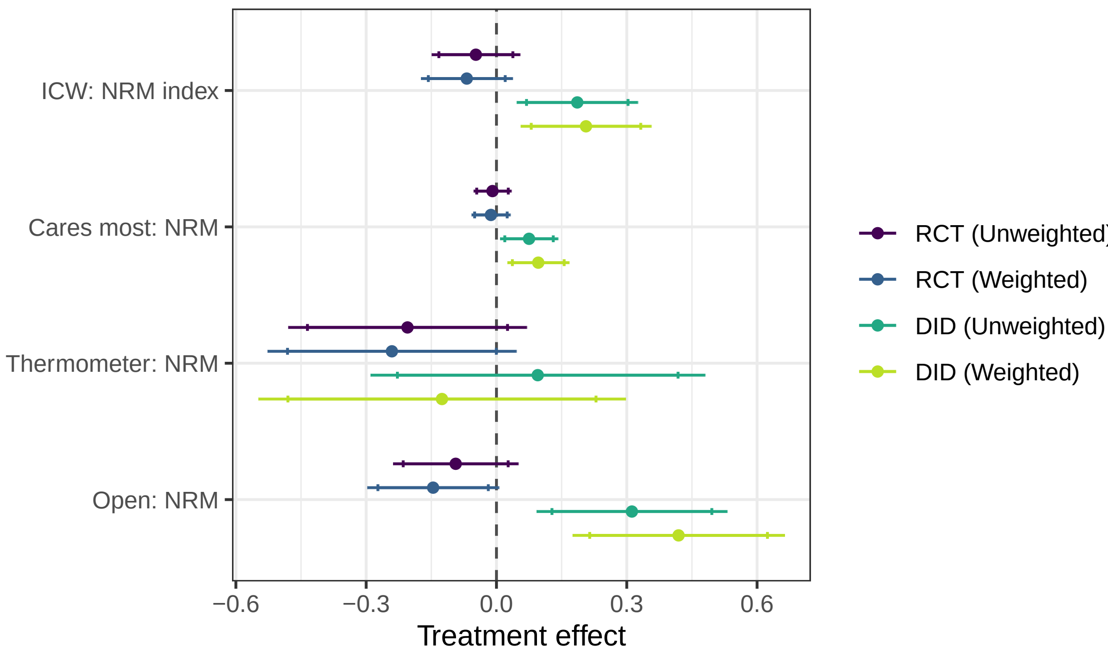


Figure A15: Reweighted estimates from experimental and difference-in-differences research designs

Notes: The RCT specifications are estimated using equation (1); the DID specifications are estimated using equation 2. We vary the inclusion of inverse weights based on a participant's predicted increase in social media usage due to the subsidy treatment (RCT) or VPN access during social media ban (DID) (see Figure A10). 90% and 95% confidence intervals plotted (two-sided tests).

A.8 Additional Tables

Table A7: Comparison of samples used across research designs

	A. Baseline sample			**B. Field experiment sample**								**C. Difference-in-differences sample**							
					All		Audited		Not audited				All		Audited		Not audited		
	N^{BL}	μ	σ^2	N^{RCT}	μ	σ^2	μ_{WA}	σ^2_{WA}	$\mu_{WA'}$	$\sigma^2_{WA'}$	p(WA=WA')	N^{DID}	μ	σ^2	μ_{WA}	σ^2_{WA}	$\mu_{WA'}$	$\sigma^2_{WA'}$	p(WA=WA')
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Age	1542	30.45	8.02	1389	30.58	8.09	30.08	8.15	31.03	8.01	[0.03]	1310	30.48	8.04	30.03	8.07	30.89	8.00	[0.06]
Male	1542	0.58	0.49	1389	0.67	0.47	0.66	0.48	0.69	0.46	[0.10]	1310	0.68	0.47	0.67	0.47	0.69	0.46	[0.12]
MTN	1542	0.41	0.49	1389	0.46	0.50	0.41	0.49	0.51	0.50	[1.00]	1310	0.48	0.50	0.40	0.49	0.55	0.50	[1.00]
Primary education	1542	0.91	0.28	1389	0.91	0.28	0.92	0.28	0.91	0.28	[0.67]	1310	0.92	0.28	0.91	0.28	0.92	0.28	[0.77]
Secondary education	1542	0.55	0.50	1389	0.55	0.50	0.61	0.49	0.50	0.50	[0.00]	1310	0.56	0.50	0.61	0.49	0.52	0.50	[0.02]
Living conditions rel. to others	1542	3.32	0.83	1389	3.33	0.83	3.33	0.80	3.32	0.87	[0.58]	1310	3.32	0.81	3.36	0.80	3.28	0.81	[0.93]
Traditional Christian	1542	0.61	0.49	1389	0.60	0.49	0.57	0.50	0.62	0.48	[0.02]	1310	0.61	0.49	0.58	0.49	0.64	0.48	[0.60]
Evangelical Christian	1542	0.20	0.40	1389	0.20	0.40	0.21	0.41	0.20	0.40	[0.06]	1310	0.19	0.39	0.20	0.40	0.18	0.39	[0.17]
Muslim	1542	0.18	0.38	1389	0.19	0.39	0.21	0.41	0.17	0.38	[0.74]	1310	0.19	0.39	0.21	0.41	0.17	0.37	[0.42]
NRM supporter	1542	0.27	0.45	1389	0.28	0.45	0.24	0.43	0.31	0.46	[0.13]	1310	0.25	0.44	0.24	0.43	0.27	0.44	[0.21]
ICW: Social media use	1542	-0.01	0.97	1253	0.01	0.99	0.10	1.02	-0.10	0.95	[0.01]	1310	0.00	0.99	0.09	1.01	-0.08	0.96	[0.06]
ICW: Political knowledge	1542	-0.02	1.01	1253	0.02	1.01	-0.04	1.02	0.08	0.98	[0.09]	1310	0.01	1.01	-0.03	1.03	0.04	0.98	[0.65]
ICW: Govt support	1542	0.01	0.99	1253	-0.01	0.99	-0.05	1.00	0.05	0.98	[0.03]	1310	0.00	0.99	-0.08	1.01	0.08	0.97	[0.01]
ICW: Govt performance approval	1542	-0.01	1.01	1253	-0.01	1.01	-0.04	1.02	0.01	1.00	[0.13]	1310	-0.01	1.00	-0.01	1.00	-0.02	1.01	[0.13]
ICW: Support democracy	1542	0.00	1.00	1253	0.03	0.99	0.00	0.98	0.07	1.01	[0.17]	1310	0.03	1.00	-0.03	0.96	0.08	1.02	[0.00]
ICW: Political polarization	1542	0.01	0.94	1253	0.01	0.98	0.04	0.99	-0.02	0.97	[0.32]	1310	0.01	0.98	0.03	1.02	-0.01	0.94	[0.79]
ICW: COVID-19 knowledge	1542	0.03	0.98	1253	0.02	0.98	0.05	0.95	-0.02	1.02	[0.35]	1310	0.01	0.98	0.01	1.01	0.02	0.96	[0.34]
ICW: COVID-19 behavior	1542	-0.05	1.03	1253	-0.03	1.04	0.00	1.04	-0.06	1.04	[0.52]	1310	-0.04	1.04	0.01	1.03	-0.09	1.05	[0.86]
ICW: General welfare	1542	-0.02	1.01	1253	0.00	1.01	0.00	1.02	0.00	0.99	[0.84]	1310	0.00	1.01	0.02	1.02	-0.02	1.01	[0.09]

Notes: Table compares baseline (wave 1) sample characteristics of all participants (Panel A); those in the field experimental sample (Panel B); and those in difference-in-differences sample (Panel C). Within these samples, columns (5)-(6) and (13)-(14) provide mean and standard deviation of characteristics; columns (7)-(10) and (15)-(18) compare WhatsApp-audited subsample to non-audited subsample; columns (11) and (19) provide *p*-value testing for equivalence between these subsamples.

Table A8: Balance in field experiment

	A. Full sample						**B. NRM supporters**		**C. Non-NRM supporters**	
	Wave 1			Wave 2			Wave 1	Wave 2	Wave 1	Wave 2
	N^{BL}	μ_C	τ^{BL}	N^{ML}	μ_C	τ^{ML}	τ^{BL}	τ^{ML}	τ^{BL}	τ^{ML}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age	1389	30.78	-0.25 [0.54]				-0.83 [0.53]		0.08 [0.87]	
Male	1389	0.67	0.00 [0.84]				0.00 [0.96]		0.02 [0.44]	
MTN	1389	0.46	0.00 [1.00]				0.00 [1.00]		0.00 [1.00]	
Primary education	1389	0.92	-0.01 [0.45]				-0.08 [0.11]		0.01 [0.68]	
Secondary education	1389	0.56	-0.01 [0.63]				0.03 [0.69]		0.00 [0.93]	
Living conditions rel. to others	1389	3.36	-0.06 [0.18]				0.07 [0.56]		-0.09 [0.09]	
Traditional Christian	1389	0.59	0.01 [0.64]				0.07 [0.37]		0.02 [0.52]	
Evangelical Christian	1389	0.21	-0.01 [0.63]				0.02 [0.78]		-0.03 [0.31]	
Muslim	1389	0.19	0.00 [0.95]				-0.08 [0.22]		0.00 [0.88]	
NRM supporter	1389	0.29	-0.03 [0.18]				0.00 [0.96]		-0.03 [0.19]	
ICW: Social media use	1253	0.00	-0.05 [0.29]				-0.17 [0.37]		-0.02 [0.70]	
ICW: Political knowledge	1253	0.01	0.00 [0.97]	1389	0.01	-0.07 [0.12]	-0.07 [0.68]	-0.19 [0.07]	0.04 [0.52]	-0.06 [0.25]
ICW: Govt support	1253	0.01	-0.02 [0.67]	1389	0.01	-0.01 [0.79]	0.04 [0.77]	0.01 [0.85]	-0.03 [0.59]	0.02 [0.74]
ICW: Govt performance approval	1253	-0.01	0.01 [0.82]	1389	0.00	0.02 [0.66]	-0.08 [0.62]	0.05 [0.67]	0.01 [0.84]	-0.03 [0.66]
ICW: Support democracy	1253	0.00	0.11 [0.04]	1389	-0.01	0.03 [0.56]	0.09 [0.64]	-0.04 [0.58]	0.09 [0.14]	0.08 [0.14]
ICW: Political polarization	1253	0.00	0.05 [0.37]	1389	0.00	-0.06 [0.23]	0.10 [0.57]	0.00 [0.97]	0.08 [0.22]	0.00 [0.99]
ICW: COVID-19 knowledge	1253	0.01	0.03 [0.51]	1389	-0.01	0.05 [0.34]	-0.04 [0.76]	0.23 [0.13]	-0.01 [0.91]	-0.01 [0.87]
ICW: COVID-19 behavior	1253	0.01	-0.08 [0.14]	1389	0.01	-0.01 [0.77]	-0.21 [0.16]	0.03 [0.83]	-0.07 [0.24]	-0.03 [0.68]
ICW: General welfare	1253	0.01	-0.02 [0.70]	1389	0.02	-0.09 [0.05]	0.07 [0.63]	-0.10 [0.39]	-0.05 [0.41]	-0.07 [0.19]

Notes: Each specification is estimated using OLS including block and enumerator fixed effects using measures defined both in wave 1 (columns 1-3) and wave 2 (columns 4-6). Columns (3) and (6) provide coefficients testing for differences with heteroskedasticity-robust p-values in square brackets. Panels B and C provide equivalent tests when restricting sample to NRM supporter (or not) as defined in wave 2.

Table A9: Tests of attrition in the field experiment

	Outcome: Attrited			
	(1)	(2)	(3)	(4)
Treatment	0.003	0.005	0.007	0.014
	(0.011)	(0.011)	(0.014)	(0.014)
NRM supporter			-0.014	0.007
			(0.016)	(0.007)
Treatment $\times$ NRM supporter			-0.020	-0.026
			(0.021)	(0.024)
Treatment + Treatment $\times$ NRM supporter			-0.012	-0.011
			(0.017)	(0.020)
Observations	1,455	1,455	1,455	1,455
Control mean	0.04	0.04	0.04	0.04
Control SD	0.21	0.21	0.21	0.21
Block FEs		✓		✓

Notes: Each specification is estimated using OLS, with even-indexed columns adding randomization block fixed effects. The sum of coefficients reports the treatment effect among NRM supporting participants. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Robustness of partisan-moderated effects of experimental social media subsidy on NRM support, by outcome index construction

	Support for NRM				**Support for opposition**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Z-score index of support								
Treatment	-0.047	-0.053	-0.005	-0.009	0.021	0.024	-0.014	-0.016
	(0.052)	(0.052)	(0.066)	(0.067)	(0.053)	(0.053)	(0.068)	(0.068)
Treatment × NRM supporter			-0.306**	-0.302**			0.193	0.118
			(0.121)	(0.132)			(0.133)	(0.147)
Treatment + Treatment × NRM supporter			-0.311***	-0.311***			0.178	0.102
			(0.101)	(0.113)			(0.115)	(0.131)
Observations	1,253	1,253	1,253	1,253	1,253	1,253	1,253	1,253
Control mean	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00
Control SD	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01
LASSO-selected covariates		✓		✓		✓		✓
Panel B: First principal component of support								
Treatment	-0.082	-0.110*	-0.020	-0.047	0.025	0.032	-0.026	-0.022
	(0.068)	(0.066)	(0.088)	(0.087)	(0.064)	(0.064)	(0.084)	(0.083)
Treatment × NRM supporter			-0.365**	-0.318*			0.281*	0.276*
			(0.156)	(0.167)			(0.156)	(0.159)
Treatment + Treatment × NRM supporter			-0.385***	-0.365**			0.255*	0.254*
			(0.129)	(0.143)			(0.131)	(0.136)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,387
Control mean	0.04	0.04	0.04	0.04	-0.02	-0.02	-0.02	-0.02
Control SD	1.37	1.37	1.37	1.37	1.28	1.28	1.28	1.28
LASSO-selected covariates		✓		✓		✓		✓
Panel C: ICW difference between NRM and Opposition index of support								
Treatment	-0.045	-0.057	-0.010	-0.016				
	(0.050)	(0.050)	(0.065)	(0.062)				
Treatment × NRM supporter			-0.248**	-0.206*				
			(0.117)	(0.118)				
Treatment + Treatment × NRM supporter			-0.258***	-0.221**				
			(0.098)	(0.100)				
Observations	1,389	1,389	1,389	1,387				
Control mean	-0.00	-0.00	-0.00	-0.00				
Control SD	1.00	1.00	1.00	1.00				
LASSO-selected covariates		✓		✓				

Notes: Each specification is estimated using OLS, and adjusts for the baseline and midline pre-treatment outcome and block and wave 3 enumerator fixed effects. Even-indexed columns add LASSO-selected covariates. To estimate heterogeneous treatment effects, the indicator for NRM supporter is fully interacted with treatment indicator and fixed effects (and LASSO-selected covariates, when relevant). Lower-order terms are omitted to save space. The sum of coefficients reports the treatment effect among NRM supporting participants. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Field experimental test of spillovers

	Support for NRM				Support for opposition			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Support for NRM and opposition party ICW indexes								
Number treated in TC	-0.014	-0.026	-0.033	-0.052	-0.008	0.000	-0.001	-0.050
	(0.018)	(0.019)	(0.025)	(0.051)	(0.019)	(0.020)	(0.028)	(0.058)
Number treated in TC × NRM supporter			0.034	-0.014			-0.009	0.067
			(0.053)	(0.117)			(0.047)	(0.156)
N. treated in TC + N. treated in TC × NRM supporter			0.002	-0.066			-0.011	0.017
			(0.046)	(0.106)			(0.037)	(0.145)
Observations	696	696	696	696	696	696	696	696
Control mean	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00
Control SD	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO-selected covariates		✓		✓		✓		✓
Panel B: Which party cares most about people like the respondent								
Number treated in TC	-0.001	-0.003	-0.006	-0.004	0.003	0.004	0.003	0.010
	(0.008)	(0.008)	(0.013)	(0.023)	(0.007)	(0.007)	(0.011)	(0.023)
Number treated in TC × NRM supporter			0.022	-0.062			-0.005	0.038
			(0.025)	(0.073)			(0.018)	(0.051)
N. treated in TC + N. treated in TC × NRM supporter			0.015	-0.065			-0.001	0.048
			(0.021)	(0.069)			(0.014)	(0.046)
Observations	696	696	696	696	696	696	696	696
Control mean	0.72	0.72	0.72	0.72	0.24	0.24	0.24	0.24
Control SD	0.45	0.45	0.45	0.45	0.43	0.43	0.43	0.43
LASSO-selected covariates		✓		✓		✓		✓
Panel C: Feeling thermometer (0-very cold – 10-very warm)								
Number treated in TC	0.016	-0.005	0.012	0.049	0.068	0.062	0.131*	0.044
	(0.052)	(0.052)	(0.075)	(0.140)	(0.048)	(0.050)	(0.071)	(0.146)
Number treated in TC × NRM supporter			-0.072	0.014			-0.123	0.057
			(0.125)	(0.385)			(0.141)	(0.480)
N. treated in TC + N. treated in TC × NRM supporter			-0.060	0.064			0.008	0.101
			(0.100)	(0.359)			(0.122)	(0.457)
Observations	696	696	696	696	696	696	696	696
Control mean	5.96	5.96	5.96	5.96	5.41	5.41	5.41	5.41
Control SD	2.79	2.79	2.79	2.79	2.64	2.64	2.64	2.64
LASSO-selected covariates		✓		✓		✓		✓
Panel D: Openness to voting for party (1-not at all – 5-very open)								
Number treated in TC	-0.048*	-0.064**	-0.093**	-0.180**	-0.063**	-0.049	-0.072	-0.120
	(0.028)	(0.029)	(0.043)	(0.085)	(0.030)	(0.031)	(0.047)	(0.084)
Number treated in TC × NRM supporter			0.089	0.056			0.055	0.198
			(0.072)	(0.811)			(0.075)	(0.307)
N. treated in TC + N. treated in TC × NRM supporter			-0.004	-0.125			-0.018	0.079
			(0.057)	(0.807)			(0.059)	(0.295)
Observations	696	696	696	696	696	696	696	696
Control mean	3.35	3.35	3.35	3.35	3.08	3.08	3.08	3.08
Control SD	1.44	1.44	1.44	1.44	1.43	1.43	1.43	1.43
LASSO-selected covariates		✓		✓		✓		✓

Notes: Each specification is estimated using OLS, and adjusts for the baseline and midline pre-treatment outcome and block and wave 3 enumerator fixed effects. The sample is restricted to respondents assigned to control. "Number treated in TC" indicates the total number of treated respondents in the same trading center as a given respondent; the total number of respondents overall is also adjusted for. To estimate heterogeneous effects, the indicator for NRM supporter is fully interacted with number treated in TC, fixed effects, and LASSO-selected covariates. To estimate heterogeneous effects, the indicator for NRM supporter is fully interacted with number treated in TC, fixed effects, and LASSO-selected covariates. Lower-order terms are omitted to save space. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Effects of experimental social media subsidy on perceptions of survey enumerators or study purpose

	Outcome: Varies by panel															
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Panel A: Believe enumerators were sent by...																
	Research organization				**Government/NRM**				**NGO**				**Other**			
Treatment	-0.00	-0.00	0.01	0.01	-0.00	-0.00	-0.00	-0.01	-0.01	-0.01	-0.01	-0.00	0.01	0.00	0.00	-0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Treatment × NRM supporter			-0.01	0.00			0.01	0.01			-0.00	-0.02			0.01	0.01
			(0.05)	(0.05)			(0.03)	(0.03)			(0.03)	(0.03)			(0.03)	(0.03)
Treatment + Treatment × NRM supporter			-0.001	0.008			0.010	0.008			-0.014	-0.027			0.008	0.004
			(0.041)	(0.042)			(0.024)	(0.025)			(0.031)	(0.032)			(0.022)	(0.024)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,387	1,389	1,389
Control mean	0.88	0.88	0.88	0.88	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Control SD	0.33	0.33	0.33	0.33	0.21	0.21	0.21	0.21	0.19	0.19	0.19	0.19	0.20	0.20	0.20	0.20
LASSO-selected covariates		✓		✓		✓		✓		✓		✓		✓		✓
Panel B: Purpose of study																
	Study social media effects on society				**Help government to monitor citizens**				**Study support for government**				**Study citizen beliefs about COVID-19**			
Treatment	-0.01	-0.01	0.01	0.01	0.00	-0.00	0.01	0.01	0.00	-0.00	-0.00	-0.01	0.01	0.01	-0.01	-0.00
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Treatment × NRM supporter			-0.04	-0.07			0.04	0.05			0.05	0.07**			-0.06	-0.06
			(0.07)	(0.07)			(0.04)	(0.04)			(0.03)	(0.03)			(0.07)	(0.07)
Treatment + Treatment × NRM supporter			-0.032	-0.054			0.051	0.051			0.043	0.055*			-0.067	-0.061
			(0.063)	(0.063)			(0.038)	(0.038)			(0.030)	(0.029)			(0.063)	(0.064)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389
Control mean	0.32	0.32	0.32	0.32	0.14	0.14	0.14	0.14	0.08	0.08	0.08	0.08	0.43	0.43	0.43	0.43
Control SD	0.47	0.47	0.47	0.47	0.35	0.35	0.35	0.35	0.27	0.27	0.27	0.27	0.49	0.49	0.49	0.49
LASSO-selected covariates		✓		✓		✓		✓		✓		✓		✓		✓

Notes: Each specification is estimated using OLS, and adjusts for the baseline and midline pre-treatment outcome (except in Panel B, since it was not asked) and block and wave 3 enumerator fixed effects. Even-indexed columns add LASSO-selected covariates. To estimate heterogeneous treatment effects, the indicator for NRM supporter is fully interacted with treatment indicator and fixed effects (and LASSO-selected covariates, when relevant). Lower-order terms are omitted to save space. The sum of coefficients reports the treatment effect among NRM supporting participants. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: Experimental treatment effects on party support, by prior VPN usage

	Full sample				NRM supporters				Non-NRM supporters			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Support for NRM and opposition party ICW indexes												
	NRM		**Opposition**		**NRM**		**Opposition**		**NRM**		**Opposition**	
Treatment	-0.180	-0.003	0.100	-0.034	-0.382	-0.468***	0.239	0.295	-0.206	0.047	0.114	-0.037
	(0.117)	(0.064)	(0.112)	(0.066)	(0.430)	(0.173)	(0.359)	(0.189)	(0.181)	(0.081)	(0.174)	(0.081)
Observations	424	965	424	965	123	258	123	258	301	707	301	707
Control mean	0.07	-0.03	-0.02	0.01	0.34	0.26	-0.32	-0.29	-0.04	-0.14	0.11	0.12
Control SD	0.99	1.00	0.95	1.02	0.90	0.95	0.90	0.98	1.00	1.00	0.95	1.01
VPN user	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓
Panel B: Which party cares most about people like the respondent												
	NRM		**Opposition**		**NRM**		**Opposition**		**NRM**		**Opposition**	
Treatment	-0.039	0.013	0.026	0.013	-0.108	-0.163**	0.070	0.135*	-0.100	0.033	0.066	0.002
	(0.047)	(0.029)	(0.041)	(0.028)	(0.121)	(0.076)	(0.100)	(0.073)	(0.080)	(0.037)	(0.070)	(0.036)
Observations	424	965	424	965	123	258	123	258	301	707	301	707
Control mean	0.75	0.71	0.22	0.25	0.86	0.81	0.12	0.16	0.71	0.66	0.27	0.29
Control SD	0.43	0.46	0.42	0.44	0.35	0.39	0.33	0.36	0.46	0.47	0.44	0.46
VPN user	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓
Panel C: Feeling thermometer (0-very cold – 10-very warm)												
	NRM		**Opposition**		**NRM**		**Opposition**		**NRM**		**Opposition**	
Treatment	-0.549	-0.013	0.050	-0.095	-0.439	-0.974**	1.104	0.868*	-0.505	-0.018	-0.004	-0.256
	(0.335)	(0.176)	(0.300)	(0.170)	(1.439)	(0.465)	(1.160)	(0.457)	(0.479)	(0.225)	(0.435)	(0.204)
Observations	424	965	424	965	123	258	123	258	301	707	301	707
Control mean	6.09	5.89	5.44	5.40	6.71	6.77	4.75	4.53	5.82	5.56	5.75	5.74
Control SD	2.78	2.79	2.59	2.67	2.85	2.56	2.44	2.59	2.72	2.80	2.60	2.62
VPN user	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓
Panel D: Openness to voting for different party (1-not at all – 5-very open)												
	NRM		**Opposition**		**NRM**		**Opposition**		**NRM**		**Opposition**	
Treatment	-0.272	-0.057	0.195	-0.100	-0.626	-0.466*	-0.013	0.133	-0.215	0.044	0.181	-0.049
	(0.189)	(0.094)	(0.180)	(0.091)	(0.435)	(0.239)	(0.896)	(0.285)	(0.275)	(0.121)	(0.261)	(0.112)
Observations	424	965	424	965	123	258	123	258	301	707	301	707
Control mean	3.42	3.32	3.08	3.09	3.69	3.54	2.82	2.89	3.31	3.23	3.19	3.16
Control SD	1.46	1.42	1.42	1.44	1.33	1.35	1.43	1.41	1.51	1.44	1.41	1.45
VPN user	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓

Notes: Each specification is estimated using OLS, and adjusts for the baseline and midline pre-treatment outcome and block and wave 3 enumerator fixed effects. Sample split between participants who did not use a VPN on any day in the week prior to wave 2 enumeration (odd columns) and those who used a VPN on at least one day (even columns). Columns (1)-(4) estimate effects using full sample; Columns (5)-(8) using NRM supporters; Columns (9)-(12) using non-NRM supporters. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A14: Field experimental treatment effects, by prior beliefs and knowledge

	NRM support index (ICW)			
	(1)	(2)	(3)	(4)
A. Beliefs and exposure to perspectives				
Treatment	-0.039	-0.087	-0.046	-0.008
	(0.075)	(0.095)	(0.058)	(0.075)
Treatment × Good incumbent performance prior	-0.019	0.200		
	(0.109)	(0.158)		
Treatment × NRM supporter		0.188		-0.266**
		(0.291)		(0.133)
Treatment × NRM supporter × Good incumbent performance prior		-0.621*		
		(0.345)		
Treatment × Never saw opposition point of view prior			-0.296	0.837***
			(0.229)	(0.199)
Treatment × NRM supporter × Never saw opposition point of view prior				-0.841**
				(0.409)
Observations	1,389	1,389	1,389	1,389
Control mean	1.00	1.00	1.00	1.00
Control SD	0.00	0.00	0.00	0.00
B. Political knowledge				
Treatment	0.010	-0.101	0.017	0.019
	(0.140)	(0.209)	(0.087)	(0.116)
Treatment × Knowledge of recent political events	-0.142	0.061		
	(0.309)	(0.450)		
Treatment × NRM supporter		0.436		-0.441
		(0.424)		(0.311)
Treatment × NRM supporter × Knowledge of recent political events		-1.252		
		(0.895)		
Treatment × Knowledge of political leaders			-0.155	-0.080
			(0.145)	(0.198)
Treatment × NRM supporter × Knowledge of political leaders				0.115
				(0.457)
Observations	1,389	1,389	1,389	1,389
Control mean	1.00	1.00	1.00	1.00
Control SD	0.00	0.00	0.00	0.00

Notes: Each specification is estimated using OLS, and adjusts for the baseline and midline pre-treatment outcome and block and wave 3 enumerator fixed effects. To estimate heterogeneous treatment effects, the indicator for NRM supporter is fully interacted with treatment indicator and fixed effects. Lower-order terms are omitted to save space. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A15: Field experimental treatment effects, without imputing outcome variable

	Support for NRM				Support for opposition			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Support for NRM and opposition party ICW indexes								
Treatment	-0.057	-0.072	-0.009	0.006	0.014	0.038	-0.016	-0.002
	(0.050)	(0.048)	(0.064)	(0.063)	(0.051)	(0.050)	(0.066)	(0.066)
Treatment × NRM supporter			-0.262**	-0.283**			0.204	0.118
			(0.113)	(0.115)			(0.127)	(0.127)
Treatment + Treatment × NRM supporter			-0.271***	-0.277***			0.188*	0.116
			(0.094)	(0.096)			(0.108)	(0.111)
Observations	1,381	1,381	1,381	1,381	1,381	1,381	1,381	1,381
Control mean	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00
Control SD	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO-selected covariates		✓		✓		✓		✓
Panel B: Which party cares most about people like the respondent								
Treatment	-0.008	-0.008	0.005	0.005	0.027	0.027	0.013	0.013
	(0.022)	(0.022)	(0.030)	(0.030)	(0.021)	(0.021)	(0.029)	(0.029)
Treatment × NRM supporter			-0.086*	-0.086*			0.077	0.077
			(0.050)	(0.050)			(0.047)	(0.047)
Treatment + Treatment × NRM supporter			-0.080**	-0.080**			0.090**	0.090**
			(0.040)	(0.040)			(0.037)	(0.037)
Observations	1,383	1,383	1,383	1,383	1,383	1,383	1,383	1,383
Control mean	0.72	0.72	0.72	0.72	0.25	0.25	0.25	0.25
Control SD	0.45	0.45	0.45	0.45	0.43	0.43	0.43	0.43
LASSO-selected covariates		✓		✓		✓		✓
Panel C: Feeling thermometer (0-very cold – 10-very warm)								
Treatment	-0.194	-0.239*	-0.106	-0.140	-0.069	-0.098	-0.255	-0.200
	(0.140)	(0.140)	(0.180)	(0.185)	(0.133)	(0.131)	(0.167)	(0.165)
Treatment × NRM supporter			-0.616*	-0.717*			0.602*	0.452
			(0.344)	(0.381)			(0.329)	(0.358)
Treatment + Treatment × NRM supporter			-0.723**	-0.857**			0.347	0.252
			(0.293)	(0.333)			(0.283)	(0.324)
Observations	1,387	1,387	1,387	1,387	1,387	1,387	1,387	1,387
Control mean	5.96	5.96	5.96	5.96	5.41	5.41	5.41	5.41
Control SD	2.79	2.79	2.79	2.79	2.64	2.64	2.64	2.64
LASSO-selected covariates		✓		✓		✓		✓
Panel D: Openness to voting for party (1-not at all – 5-very open)								
Treatment	-0.092	-0.101	-0.031	-0.064	-0.002	0.064	0.010	0.032
	(0.074)	(0.071)	(0.095)	(0.097)	(0.071)	(0.069)	(0.091)	(0.091)
Treatment × NRM supporter			-0.227	-0.217			0.134	0.143
			(0.165)	(0.174)			(0.190)	(0.188)
Treatment + Treatment × NRM supporter			-0.258*	-0.281*			0.144	0.174
			(0.135)	(0.144)			(0.166)	(0.164)
Observations	1,387	1,387	1,387	1,387	1,387	1,387	1,387	1,387
Control mean	3.35	3.35	3.35	3.35	3.08	3.08	3.08	3.08
Control SD	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44
LASSO-selected covariates		✓		✓		✓		✓

Notes: Each specification is estimated using OLS, and adjusts for the baseline and midline pre-treatment outcome and block and wave 3 enumerator fixed effects. Even-indexed columns add LASSO-selected covariates. To estimate heterogeneous treatment effects, the indicator for NRM supporter is fully interacted with treatment indicator and fixed effects (and LASSO-selected covariates, when relevant). Lower-order terms are omitted to save space. The sum of coefficients reports the treatment effect among NRM supporting participants. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A16: Field experimental treatment effects, using varied size of randomization block fixed effects

	Support for NRM				Support for opposition			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: No randomization block fixed effects								
Treatment	-0.054	-0.073	-0.006	-0.035	0.015	0.056	-0.016	0.013
	(0.048)	(0.047)	(0.058)	(0.058)	(0.049)	(0.047)	(0.059)	(0.059)
Treatment × NRM supporter			-0.217**	-0.184*			0.168	0.188*
			(0.101)	(0.107)			(0.110)	(0.108)
Treatment + Treatment × NRM supporter			-0.223***	-0.219**			0.152	0.202**
			(0.083)	(0.090)			(0.093)	(0.091)
Observations	1,389	1,387	1,389	1,387	1,389	1,389	1,389	1,387
Control mean	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00
Control SD	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO-selected covariates		✓		✓		✓		✓
Panel B: Size 8 randomization block fixed effects (default)								
Treatment	-0.053	-0.080*	-0.014	-0.007	0.025	0.053	-0.018	-0.000
	(0.053)	(0.048)	(0.064)	(0.064)	(0.054)	(0.049)	(0.065)	(0.065)
Treatment × NRM supporter			-0.262**	-0.269**			0.207	0.197
			(0.113)	(0.119)			(0.126)	(0.130)
Treatment + Treatment × NRM supporter			-0.276***	-0.276***			0.189*	0.197*
			(0.093)	(0.101)			(0.108)	(0.112)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,387
Control mean	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00
Control SD	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO-selected covariates		✓		✓		✓		✓
Panel C: Size 4 randomization block fixed effects								
Treatment	-0.052	-0.085*	0.016	0.020	0.038	0.038	-0.044	-0.037
	(0.055)	(0.049)	(0.071)	(0.070)	(0.056)	(0.052)	(0.073)	(0.072)
Treatment × NRM supporter			-0.338**	-0.407***			0.381**	0.380**
			(0.148)	(0.138)			(0.160)	(0.156)
Treatment + Treatment × NRM supporter			-0.322**	-0.387***			0.337**	0.343**
			(0.130)	(0.119)			(0.142)	(0.139)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389
Control mean	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00
Control SD	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO-selected covariates		✓		✓		✓		✓
Panel D: Size 2 randomization block fixed effects								
Treatment	-0.045	-0.062	0.035	0.086	0.036	0.024	-0.095	-0.078
	(0.065)	(0.064)	(0.095)	(0.094)	(0.068)	(0.066)	(0.102)	(0.098)
Treatment × NRM supporter			-0.510*	-0.568**			0.696***	0.501*
			(0.284)	(0.288)			(0.254)	(0.288)
Treatment + Treatment × NRM supporter			-0.476*	-0.481*			0.601**	0.423
			(0.268)	(0.272)			(0.233)	(0.271)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,387
Control mean	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00
Control SD	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO-selected covariates		✓		✓		✓		✓

Notes: Each specification is estimated using OLS, and adjusts for the baseline and midline pre-treatment outcome and block and wave 3 enumerator fixed effects. Even-indexed columns add LASSO-selected covariates. To estimate heterogeneous treatment effects, the indicator for NRM supporter is fully interacted with treatment indicator and fixed effects (and LASSO-selected covariates, when relevant). Lower-order terms are omitted to save space. The sum of coefficients reports the treatment effect among NRM supporting participants. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: Field experimental treatment effects on self-reported media consumption behaviors

	Any hours consumed				Log+1 hours consumed			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Facebook								
Treatment	0.050***	0.056***	0.056**	0.058**	0.126**	0.117**	0.119*	0.099
	(0.019)	(0.018)	(0.024)	(0.023)	(0.053)	(0.051)	(0.066)	(0.063)
Treatment × NRM supporter			-0.040	-0.037			-0.160	-0.175
			(0.048)	(0.049)			(0.138)	(0.133)
Treatment + Treatment × NRM supporter			0.016	0.022			-0.042	-0.076
			(0.042)	(0.044)			(0.122)	(0.119)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389
Control mean	0.81	0.81	0.81	0.81	1.94	1.94	1.94	1.94
Control SD	0.39	0.39	0.39	0.39	1.16	1.16	1.16	1.16
LASSO-selected covariates		✓		✓		✓		✓
Panel B: WhatsApp								
Treatment	0.028*	0.033**	0.031*	0.034*	0.087*	0.070	0.076	0.055
	(0.015)	(0.014)	(0.018)	(0.018)	(0.048)	(0.045)	(0.058)	(0.056)
Treatment × NRM supporter			-0.036	-0.029			-0.099	-0.063
			(0.037)	(0.037)			(0.129)	(0.127)
Treatment + Treatment × NRM supporter			-0.005	0.006			-0.023	-0.008
			(0.033)	(0.033)			(0.115)	(0.114)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389
Control mean	0.89	0.89	0.89	0.89	2.26	2.26	2.26	2.26
Control SD	0.31	0.31	0.31	0.31	1.07	1.07	1.07	1.07
LASSO-selected covariates		✓		✓		✓		✓
Panel C: Facebook Messenger								
Treatment	0.045*	0.043*	0.060*	0.056*	0.088*	0.065	0.086	0.066
	(0.024)	(0.023)	(0.031)	(0.030)	(0.052)	(0.050)	(0.065)	(0.064)
Treatment × NRM supporter			-0.072	-0.063			-0.198	-0.153
			(0.066)	(0.065)			(0.144)	(0.137)
Treatment + Treatment × NRM supporter			-0.013	-0.006			-0.113	-0.087
			(0.058)	(0.057)			(0.129)	(0.122)
Observations	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389
Control mean	0.62	0.62	0.62	0.62	1.26	1.26	1.26	1.26
Control SD	0.49	0.49	0.49	0.49	1.21	1.21	1.21	1.21
LASSO-selected covariates		✓		✓		✓		✓
Panel D: Twitter								
Treatment	0.028*	0.035**	0.037*	0.035*	0.053*	0.062**	0.064*	0.059*
	(0.016)	(0.016)	(0.021)	(0.021)	(0.027)	(0.027)	(0.036)	(0.035)
Treatment × NRM supporter			-0.050	-0.044			-0.067	-0.033
			(0.044)	(0.043)			(0.078)	(0.076)
Treatment + Treatment × NRM supporter			-0.013	-0.009			-0.004	0.026
			(0.038)	(0.038)			(0.069)	(0.067)
Observations	1,389	1,387	1,389	1,389	1,389	1,389	1,389	1,389
Control mean	0.09	0.09	0.09	0.09	0.15	0.15	0.15	0.15
Control SD	0.29	0.29	0.29	0.29	0.49	0.49	0.49	0.49
LASSO-selected covariates		✓		✓		✓		✓

Notes: Each specification is estimated using OLS, and adjusts for the baseline and midline pre-treatment outcome and block and wave 3 enumerator fixed effects. Even-indexed columns add LASSO-selected covariates. To estimate heterogeneous treatment effects, the indicator for NRM supporter is fully interacted with treatment indicator and fixed effects (and LASSO-selected covariates, when relevant). Lower-order terms are omitted to save space. The sum of coefficients reports the treatment effect among NRM supporting participants. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A18: Balance in difference-in-differences design

	N^{BL+ML}	$\mu_{VPN=0}$	$\sigma^2_{VPN=0}$	τ^1	τ^2	τ^3
	(1)	(2)	(3)	(4)	(5)	(6)
Age	1310	31.40	8.18	-2.12 [0.00]	-1.48 [0.00]	
Male	1310	0.70	0.46	-0.01 [0.64]	-0.04 [0.26]	-0.03 [0.31]
MTN	1310	0.54	0.50	0.01 [0.65]	0.01 [0.65]	0.02 [0.25]
Primary education	1310	0.91	0.28	0.02 [0.22]	0.02 [0.21]	0.02 [0.26]
Secondary education	1310	0.57	0.50	-0.03 [0.32]	-0.01 [0.64]	-0.02 [0.55]
Living conditions rel. to others	1310	3.33	0.77	-0.01 [0.75]	-0.02 [0.72]	-0.01 [0.90]
Traditional Christian	1310	0.62	0.49	-0.01 [0.66]	0.00 [0.98]	0.01 [0.85]
Evangelical Christian	1310	0.19	0.39	0.01 [0.75]	0.02 [0.43]	0.02 [0.55]
Muslim	1310	0.16	0.37	0.03 [0.34]	0.00 [0.99]	0.00 [0.95]
NRM supporter	1310	0.29	0.46	-0.03 [0.28]	-0.02 [0.53]	-0.02 [0.62]
ICW: Social media use	1310	-0.10	1.01	0.11 [0.07]	0.09 [0.17]	0.07 [0.30]
ICW: Political knowledge	1310	-0.04	1.04	-0.06 [0.19]	-0.04 [0.34]	-0.01 [0.90]
ICW: Govt support	1310	0.14	0.97	-0.16 [0.01]	-0.14 [0.04]	-0.12 [0.08]
ICW: Govt performance approval	1310	0.06	0.95	-0.10 [0.08]	-0.09 [0.16]	-0.07 [0.26]
ICW: Support democracy	1310	0.03	0.95	-0.04 [0.52]	0.00 [0.96]	0.02 [0.77]
ICW: Political polarization	1310	0.02	1.00	-0.01 [0.92]	-0.02 [0.70]	-0.01 [0.83]
ICW: COVID-19 knowledge	1310	0.05	1.02	-0.02 [0.65]	0.00 [0.95]	0.00 [0.97]
ICW: COVID-19 behavior	1310	-0.07	1.03	-0.01 [0.80]	-0.02 [0.71]	-0.01 [0.86]
ICW: General welfare	1310	0.00	0.99	0.02 [0.69]	0.02 [0.78]	0.02 [0.78]

Notes: The first three columns report summary statistics for our sample of baseline and midline respondents, with columns (2) and (3) restricting attention to non-VPN users. Columns (4)-(6) estimate the difference between VPN and non-VPN users by regressing wave 1-defined variables on an indicator for baseline VPN usage; trading center-clustered p-values are in brackets. Column (4) uses enumerator fixed effects; column (5) additionally adds trading center fixed effects; column (6) additionally adds age fixed effects.

Table A19: Differences in midline attrition by baseline VPN use

	Outcome: Attrited	
	(1)	(2)
VPN	-0.022	-0.001
	(0.020)	(0.020)
Observations	1,542	1,538
Control mean	0.16	0.16
Control SD	0.37	0.37
Block FEs		✓

Notes: Each specification is estimated using OLS and includes the full sample of respondents that completed the baseline survey. Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A20: Differential effects of VPN use on daily WhatsApp usage during the social media ban, by baseline partisanship

	Outcome: Varies by panel					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Used Whatsapp						
VPN × WhatsApp ban	0.054***	0.047***	0.046**	0.040**	0.028	0.021
	(0.014)	(0.017)	(0.018)	(0.017)	(0.021)	(0.022)
VPN × WhatsApp ban × NRM supporter				0.050	0.040	0.033
				(0.035)	(0.042)	(0.043)
VPN × Ban + VPN × Ban × NRM supporter				0.090***	0.068*	0.055
				(0.029)	(0.035)	(0.038)
Observations	112,943	111,655	111,655	112,943	106,076	105,075
Clusters (TCs)	125	116	116	125	111	111
Control mean	0.31	0.30	0.30	0.31	0.31	0.31
Control SD	0.46	0.46	0.46	0.46	0.46	0.46
Interactive FEs		TC	TC & Age		TC	TC & Age
Panel B: Number of audited times seen on WhatsApp						
VPN × WhatsApp ban	0.084***	0.077**	0.076**	0.054	0.037	0.029
	(0.028)	(0.034)	(0.033)	(0.034)	(0.042)	(0.041)
VPN × WhatsApp ban × NRM supporter				0.105	0.070	0.052
				(0.068)	(0.081)	(0.083)
VPN × Ban + VPN × Ban × NRM supporter				0.159***	0.107	0.081
				(0.057)	(0.064)	(0.072)
Observations	112,943	111,655	111,655	112,943	106,076	105,075
Clusters (TCs)	125	116	116	125	111	111
Control mean	0.54	0.53	0.53	0.54	0.55	0.55
Control SD	0.95	0.95	0.95	0.95	0.96	0.96
Interactive FEs		TC	TC & Age		TC	TC & Age

Notes: Each specification is estimated using OLS, and includes individual and period fixed effects. To estimate heterogeneous treatment effects, the indicator for NRM supporter is fully interacted with the individual and period fixed effects. The sum of coefficients reports the difference-in-differences estimate among NRM supporting participants. Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A21: Differential effects of VPN use on support for the NRM after the social media ban, by baseline partisanship

	Support for NRM				Support for opposition			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Support for NRM and opposition party ICW indexes								
VPN × Post election	0.184**	0.196***	0.151*	0.219**	-0.114*	-0.156**	-0.053	-0.121
	(0.072)	(0.074)	(0.088)	(0.094)	(0.069)	(0.070)	(0.083)	(0.093)
VPN × Post election × NRM supporter			-0.090	-0.177			-0.023	-0.049
			(0.135)	(0.174)			(0.150)	(0.195)
VPN × Post election + VPN × Post election × NRM supporter			0.061	0.042			-0.076	-0.170
			(0.099)	(0.124)			(0.120)	(0.163)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control mean	-0.05	-0.06	-0.05	-0.06	0.06	0.07	0.06	0.07
Control SD	1.00	1.00	1.00	1.00	0.98	0.98	0.98	0.98
Trading center × Post election FEs		✓		✓		✓		✓
Panel B: Which party cares most about people like the respondent								
VPN × Post election	0.075**	0.078**	0.070*	0.084*	-0.070**	-0.071**	-0.061	-0.072
	(0.034)	(0.035)	(0.042)	(0.045)	(0.032)	(0.033)	(0.040)	(0.044)
VPN × Post election × NRM supporter			-0.032	-0.045			-0.012	-0.014
			(0.073)	(0.094)			(0.064)	(0.077)
VPN × Post election + VPN × Post election × NRM supporter			0.038	0.039			-0.073	-0.086
			(0.058)	(0.075)			(0.046)	(0.058)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control mean	0.64	0.64	0.64	0.64	0.23	0.24	0.23	0.24
Control SD	0.48	0.48	0.48	0.48	0.42	0.42	0.42	0.42
Trading center × Post election FEs		✓		✓		✓		✓
Panel C: Feeling thermometer (0-very cold – 10-very warm)								
VPN × Post election	0.095	0.285	-0.108	0.304	-0.334*	-0.438**	-0.323	-0.546**
	(0.196)	(0.208)	(0.235)	(0.247)	(0.172)	(0.181)	(0.209)	(0.233)
VPN × Post election × NRM supporter			0.254	-0.266			0.279	0.265
			(0.387)	(0.468)			(0.425)	(0.576)
VPN × Post election + VPN × Post election × NRM supporter			0.146	0.038			-0.045	-0.281
			(0.308)	(0.389)			(0.352)	(0.521)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control mean	5.79	5.79	5.79	5.79	4.99	5.00	4.99	5.00
Control SD	2.72	2.72	2.72	2.72	2.52	2.52	2.52	2.52
Trading center × Post election FEs		✓		✓		✓		✓
Panel D: Openness to voting for different party (1-not at all – 5-very open)								
VPN × Post election	0.313***	0.258**	0.316**	0.301*	0.017	-0.044	0.154	0.102
	(0.112)	(0.121)	(0.141)	(0.161)	(0.116)	(0.122)	(0.135)	(0.159)
VPN × Post election × NRM supporter			-0.299	-0.296			-0.161	-0.220
			(0.205)	(0.253)			(0.223)	(0.299)
VPN × Post election + VPN × Post election × NRM supporter			0.017	0.005			-0.007	-0.118
			(0.139)	(0.171)			(0.185)	(0.255)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control mean	3.37	3.37	3.37	3.37	3.08	3.09	3.08	3.09
Control SD	1.43	1.43	1.43	1.43	1.45	1.45	1.45	1.45
Trading center × Post election FEs		✓		✓		✓		✓

Notes: Each specification is estimated using OLS, and includes individual and period fixed effects. To estimate heterogeneous treatment effects, the indicator for NRM supporter is fully interacted with the individual and period fixed effects. The sum of coefficients reports the difference-in-differences estimate among NRM supporting participants. Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A22: Differential effects of VPN use on support for the NRM after the social media ban, conditional on covariate × period fixed effects

	Outcome: NRM support ICW index									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Non-political support interactive covariates										
VPN × Post election	0.140**	0.170**	0.109*	0.161**	0.177**	0.194**	0.134*	0.164**	0.218*	0.272*
	(0.070)	(0.073)	(0.063)	(0.066)	(0.072)	(0.075)	(0.069)	(0.070)	(0.114)	(0.140)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	910	864
R^2	0.63	0.67	0.69	0.73	0.64	0.68	0.64	0.69	0.63	0.71
Control outcome mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Control outcome std. dev.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trading center × Post election FEs		✓		✓		✓		✓		✓
Baseline covariate × Post interaction	Age FEs		Political news consumption		Political knowledge		Govt/NRM survey		Enumerator	
Panel B: Political support interactive covariates										
VPN × Post election	0.140**	0.170**	0.109*	0.161**	0.177**	0.194**	0.134*	0.164**	0.218*	0.272*
	(0.070)	(0.073)	(0.063)	(0.066)	(0.072)	(0.075)	(0.069)	(0.070)	(0.114)	(0.140)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	910	864
R^2	0.63	0.67	0.69	0.73	0.64	0.68	0.64	0.69	0.63	0.71
Control outcome mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Control outcome std. dev.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trading center × Post election FEs		✓		✓		✓		✓		✓
Baseline covariate × Post interaction	NRM		Openness to NRM		NRM thermometer		NRM LC5 vote intent		NRM MP vote intent	

Notes: Each specification is estimated using OLS, and includes individual and period fixed effects as well as interactions between the period fixed effect and the covariate(s) listed at the foot of each regression. Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A23: Differential effects of VPN use on support for the NRM after the social media ban, using entropy balancing

	Support for NRM		Support for Opposition	
	(1)	(2)	(3)	(4)
Panel A: Support for NRM and opposition party ICW indexes				
VPN × Post election	0.196**	0.201**	-0.136*	-0.174**
	(0.077)	(0.077)	(0.072)	(0.073)
Observations	2,620	2,612	2,620	2,612
Control mean	-0.05	-0.06	0.06	0.07
Control SD	1.00	1.00	0.98	0.98
Trading center × Post election FEs		✓		✓
Panel B: Which party cares most about people like the respondent				
VPN × Post election	0.076**	0.077**	-0.077**	-0.077**
	(0.037)	(0.037)	(0.033)	(0.034)
Observations	2,620	2,612	2,620	2,612
Control mean	0.64	0.64	0.23	0.24
Control SD	0.48	0.48	0.42	0.42
Trading center × Post election FEs		✓		✓
Panel C: Feeling thermometer (0-very cold – 10-very warm)				
VPN × Post election	0.136	0.255	-0.432**	-0.514***
	(0.209)	(0.220)	(0.175)	(0.179)
Observations	2,620	2,612	2,620	2,612
Control mean	5.79	5.79	4.99	5.00
Control SD	2.72	2.72	2.52	2.52
Trading center × Post election FEs		✓		✓
Panel D: Openness to voting for party (1-not at all – 5-very open)				
VPN × Post election	0.328***	0.287**	0.019	-0.044
	(0.118)	(0.123)	(0.121)	(0.127)
Observations	2,620	2,612	2,620	2,612
Control mean	3.37	3.37	3.08	3.09
Control SD	1.43	1.43	1.45	1.45
Trading center × Post election FEs		✓		✓

Notes: Each specification is estimated using OLS and includes individual and period fixed effects. Observations are weighted by entropy balanced weights following Hainmueller and Xu (2013), where weights are constructed to balance characteristics of VPN-using and non-VPN-using samples according to age, network, gender, prior government support, prior political knowledge, and support for democracy. Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A24: Differences in beliefs over who is responsible for study

	Outcome: Believe enumerators were sent by...			
	...government		**...NRM**	
	(1)	(2)	(3)	(4)
VPN × Post election	0.006	0.006	-0.003	0.001
	(0.024)	(0.026)	(0.010)	(0.010)
Observations	2,620	2,612	2,620	2,612
R^2	0.56	0.60	0.52	0.56
Control outcome mean	0.10	0.09	0.02	0.02
Control outcome std. dev.	0.29	0.29	0.13	0.14
Trading center × Post election FEs		✓		✓

Notes: Each specification is estimated using OLS, and includes individual and period fixed effects. Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A25: Differential effects of censorship experience on support for the NRM after the social media ban

	NRM support index (ICW)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ban reduced access to reliable news × Post election	-0.175***	-0.187***								
	(0.065)	(0.070)								
Ban reduced social interactions × Post election			-0.142	-0.163*						
			(0.090)	(0.096)						
Ban affected access to entertainment × Post election					-0.063	-0.066				
					(0.086)	(0.090)				
Ban affected purchase of goods/services × Post election							-0.087	-0.075		
							(0.131)	(0.140)		
Ban interfered with business/job × Post election									0.142*	0.079
									(0.074)	(0.076)
VPN × Post election		0.200***		0.206***		0.200***		0.198***		0.200***
		(0.074)		(0.075)		(0.075)		(0.075)		(0.075)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
R^2	0.62	0.67	0.62	0.66	0.62	0.66	0.62	0.66	0.62	0.66
Ban variable mean	0.58	0.58	0.78	0.78	0.18	0.18	0.10	0.10	0.40	0.40
Trading center × Post election FEs		✓		✓		✓		✓		✓

Notes: Each specification is estimated using OLS, and includes individual and period fixed effects. Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

References

Afrobarometer. 2019. "Afrobarometer Data, Uganda, Round 8, 2019.".

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review* 110(3):629–76.

American Political Science Association. 2020. "Principles and Guidance for Human Subjects Research.".

Angrist, Joshua D. and Ivan Fernandez-Val. 2013. ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework. In *Advances in Economics and Econometrics: Volume 3, Econometrics: Tenth World Congress*. Vol. 51 Cambridge University Press p. 401.

Arias, Eric, Horacio Larreguy, John Marshall and Pablo Querubín. 2022. "Priors rule: When do malfeasance revelations help or hurt incumbent parties?" *Journal of the European Economic Association* 20(4):1433–1477.

Aronow, Peter M. and Allison Carnegie. 2013. "Beyond LATE: Estimation of the Average Treatment Effect with an Instrumental Variable." *Political Analysis* 21(4):492–506.

Belloni, Alexandre, Victor Chernozhukov and Christian Hansen. 2014. "Inference on Treatment Effects After Selection Among High-dimensional Controls." *Review of Economic Studies* 81(2):608–650.

Dell, Melissa. 2024. "Deep Learning for Economists." *arXiv preprint arXiv:2407.15339* .

Gilardi, Fabrizio, Meysam Alizadeh and Maël Kubli. 2023. "Chatgpt outperforms crowd-workers for text-annotation tasks." *arXiv preprint arXiv:2303.15056* .

Guess, Andrew M., Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow et al. 2023. "How do social media feed algorithms affect attitudes and behavior in an election campaign?" *Science* 381(6656):398–404.

Hainmueller, Jens and Yiqing Xu. 2013. "Ebalance: A Stata package for entropy balancing." *Journal of Statistical Software* 54(7):1–18.

Hotz, V. Joseph, Guido W. Imbens and Julie H. Mortimer. 2005. "Predicting the efficacy of future training programs using past experiences at other locations." *Journal of Econometrics* 125(1-2):241–270.

Mechkova, Valeriya, Daniel Pemstein, Brigitte Seim and Steven Wilson. 2022. "Measuring Internet Politics: Digital Society Project (DSP) Annual Report v4.".

Ziems, Caleb, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang and Diyi Yang. 2023. "Can Large Language Models Transform Computational Social Science?" arXiv preprint arXiv:2305.03514.