ACCESS TO SOCIAL MEDIA AND SUPPORT FOR ELECTED AUTOCRATS: FIELD EXPERIMENTAL AND OBSERVATIONAL EVIDENCE FROM UGANDA*

JEREMY BOWLES† JOHN MARSHALL‡ PIA RAFFLER§

JULY 2026

Social media can engage and inform citizens, reducing support for elected autocrats who control traditional media. But this threat also encourages such rulers to restrict social media access. In Uganda's electoral autocracy, where social media content has an anti-government slant, we estimate effects of facilitating and limiting social media access on support for the long-ruling NRM party. Experimental results show that three months of individually-subsidized social media access decreased NRM support among initial supporters. In contrast, VPN users able to maintain access during an election-time social media ban became more supportive of the NRM than individuals less able to circumvent it. The difference partly reflects the government's nationwide ban both increasing the favorability of content toward the NRM encountered by social media users and eliciting greater backlash among non-VPN users. Our findings illuminate some of the trade-offs associated with restricting social media access for incumbents seeking to maintain popular support.

*We thank Kate Casey, Gemma Dipoppa, Thad Dunning, Jeff Frieden, Asim Khwaja, Zoe Marks, George Ofosu, Dan Posner, Amanda Robinson, Arturas Rozenas, Xu Xu, David Yang, and audiences at APSA, Boston University, CAPERS, CESifo, EGAP, Harvard, the LSE-NYU Political Economy conference, MPSA, NYU, Northwestern, Princeton, Stanford, UCL, University of Virginia, 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 spread of social media has the potential to 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. 2024; Zhuravskaya, Petrova and Enikolopov 2020). Yet the implications for citizens' views of their government are ambiguous. On the one hand, social media enables the creation and dissemination of engaging content that quickly reaches users, which could expose citizens to politically-relevant information or perspectives (e.g. Diamond 2010). On the other hand, large volumes of content, easily-produced misinformation, and individual and algorithmic selection of what gets viewed may instead reinforce users' prior beliefs (e.g. Sunstein 2018). Evidence from established democracies largely supports the latter perspective, finding that social media mostly exposes users to congenial content with limited effects on political preferences (Allcott et al. 2020, 2024; Arceneaux et al. 2024; Bail et al. 2018; Guess et al. 2023; Majumdar et al. 2026; Nyhan et al. 2023; Ventura et al. forthcoming; cf. Fujiwara, Müller and Schwarz 2024; Levy 2021).

As mobile internet access accelerates across the Global South, 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 consider how social media access is likely to affect mass support for elected autocrats and the extent to which these effects are conditioned by government policies to control access.

We first theorize that social media content production and consumption patterns in electoral autocracies are more likely to expose users to unfavorable news about the regime, critical framing or commentary, and opposition messaging than in established democracies. This is because incumbents often exert significant influence over, and censorship of, traditional media outlets. By providing platforms for suppressed voices that are hard to monitor or silence, social media content can be more regime-critical than traditional media. Increasing access to social media should then, all else equal, reduce support for ruling parties in electoral autocracies, especially among incumbent supporters less likely to have previously been exposed to critical information or perspectives.

However, elected autocrats need not take unfettered access to social media as given. Early optimism that it 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 2011; Tucker et al. 2017). High-capacity autocrats can use laws and intimidation to (preemptively and retrospectively) censor expression of specific content or by particular accounts, and complement this with targeted frictions and flooding to limit attention to sensitive uncensored content (Roberts 2018). Yet such precise tools are less available to incumbents in the world's many lower-capacity electoral autocracies, which typically rely on blunter frictions like periodic internet blackouts, platform bans, and

high usage fees—and reinforce them with fear of selective sanctions. These tools have become foundational to the survival strategies of modern "informational autocrats," who control the media environment to retain popular support without resorting to repression (Guriev and Treisman 2019) or limit collective action (King, Pan and Roberts 2014; Roberts 2018).

While social media helps to coordinate anti-government protests in electoral autocracies (e.g. Castells 2015; Enikolopov, Makarin and Petrova 2020; Steinert-Threlkeld 2017), its consequences for attitudes toward ruling parties may be more subtly conditioned by government-imposed access restrictions. In addition to reducing citizens' consumption of regime-critical online content, government policies which partially restrict if, and how, citizens use social media could also influence incumbent support by altering content production and citizen welfare. First, when differentially enforced against opponents—as is typical in electoral autocracies—blunt censorship could achieve a chilling effect by reducing the pro-opposition slant or composition of content producers on social media. Second, citizens who lose access to social media may sanction the government for making platforms they value for news, social networks, entertainment, or business purposes inaccessible. These content favorability and backlash channels point to countervailing effects on incumbent support of the blunt tools typically wielded by lower-capacity autocrats to restrict social media.

This article examines how facilitating and limiting social media access affects support for Uganda's long-ruling incumbent party around the 2021 elections. A canonical electoral autocracy, the National Resistance Movement (NRM) led by President Museveni has ruled for four decades and exerts substantial control over traditional media sources. By contrast, social media is disproportionately used by opposition-leaning media and politicians—most prominently the main challenger party, the National Unity Platform (NUP). Uganda's most popular platforms, Facebook and WhatsApp, enable the production and sharing of critical news and commentary about the government as well as broader political discussion and the potential to mobilize collective action.[1] In response to these threats to the NRM's power, the government has largely deployed blunt tools based on fear and especially friction. In addition to arresting prominent critics, mass access to social media has been restricted by introducing daily fees for 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 measure the consequences for incumbent support of alleviating cost-based frictions to social media use *and* of government-imposed restrictions on access. Our field experiment subsidized access to social media for treated study participants after the 2021 elections, without altering social media use in the broader population. This identifies the persuasive effects of greater access to social

[1] As we discuss more below, as in many countries in the Global South, WhatsApp is not just a private messenger app, but also a form of mass communication via large groups which facilitate the sharing and discussion of—often highly political—content (e.g. Cheeseman et al. 2020; Ventura et al. forthcoming).

media, holding its typical regime-critical slant fixed. To instead illuminate the consequences of a nationwide policy that limited access to social media around the elections, we leverage a difference-in-differences design around Uganda's social media ban to compare citizens who were more and less able to circumvent it. Both research designs draw from an original three-wave panel survey of occasional social media users in 11 electorally competitive districts. We augment these analyses with two forms of social media data: nearly 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 provides a behavioral proxy for their social media usage in general.

The field experiment finds that increasing individuals' social media access in isolation decreased incumbent support among prior supporters. Treated respondents received mobile data and had their social media tax paid to facilitate social media use for three months following the 2021 elections; the control group instead received a smaller mobile money transfer, which could be used for any purpose. The intervention increased the likelihood of using social media on a given day, but only modestly reduced NRM support for our *average* respondent. However, NRM support significantly declined among its prior supporters, who became more supportive of opposition parties. Prior NRM supporters were more likely to encounter novel critical news or perspectives and update negatively about government performance. In contrast, non-NRM respondents' views hardly changed. This effect heterogeneity shows that greater access to social media dissipates core support for elected autocrats, conversely suggesting that financial frictions might entrench their support.

The results from the election-time social media ban reveal differential effects of blunt government policies to restrict social media access on incumbent support. Our difference-in-differences design compares changes in NRM support across respondents who did and did not already use virtual private networks (VPNs), which enabled individuals to circumvent the ban. Our behavioral data confirms that prior VPN users were more likely to use social media on a given day during the ban than non-VPN users, but—unlike the field experiment—came to view the NRM *relatively positively*. The social media ban thus led prior VPN users to become more supportive of the NRM and less supportive of opposition parties relative to non-VPN users.

The distinct effect of Uganda's social media ban appears to be driven, at least in part, by two mechanisms activated by government policies that increase fear and especially frictions of accessing social media. First, our corpus of Facebook posts shows that the social media content available to VPN users became relatively less critical of the regime after the election-time ban was imposed. Along with the natural decline in political engagement after the elections, a page-level analysis of posting behavior is consistent with the ban preventing the collection of evidence of electoral fraud and disproportionately reducing the production of politically sensitive content by regime critics. Suggesting that social media content persuaded opponents rather than calcified existing views, rel-

atively greater support for the NRM among VPN users was concentrated among respondents whose prior beliefs about NRM government performance and Ugandan democracy were least favorable. Second, respondents who lost access to social media due to the ban became relatively less supportive of the NRM. Such sanctioning is greatest among individuals who lost access to reliable news. Together, the more favorable content consumed by continuing social media users and the backlash among citizens who lost access appear to outweigh, at least in this election period, the effect of exposure to critical content established by the field experiment. While the ban surely served shorter-term purposes in disrupting collective action, these results highlight longer-term attitudinal consequences of imposing stringent barriers to mass social media access.

This article thus makes three main contributions. First, we provide experimental evidence of how greater access to social media can threaten support for elected autocrats. In established democracies where traditional media is largely free and independent, deactivating Facebook, Instagram, WhatsApp, and X suggests that social media informs users without systematically altering their political attitudes (Allcott et al. 2020, 2024; Arceneaux et al. 2024; Gauthier et al. 2026; Majumdar et al. 2026; Ventura et al. forthcoming). Our evidence from an electoral autocracy is instead consistent with social media reducing support for incumbents by exposing their supporters to counter-attitudinal content. This aligns with observational studies finding that *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). These results illustrate the greater political impact of alleviating frictions to social media where traditional media is largely under government control, though democracies are not immune to digital censorship, monitoring, and flooding (Meserve and Pemstein 2018).

Second, our results highlight effects of social media in electoral autocracies beyond collective action. By establishing that reduced incumbent support is driven by regime supporters, the depolarizing effect of social media in Uganda contrasts with the lower well-being (Allcott et al. 2020; Braghieri, Levy and Makarin 2022), reduced social cohesion (Enikolopov et al. 2024), and increased hate crime and xenophobia (Bursztyn et al. 2024; Müller and Schwarz 2023) caused by social media in advanced democracies.[2] Our findings instead point to pro-democratic effects of reducing media asymmetries via social media in elected autocracies, complementing evidence that social media can coordinate citizen protests against authoritarian governments (Enikolopov, Makarin and Petrova 2020; Qin, Strömberg and Wu 2024; Steinert-Threlkeld 2017).

Third, by highlighting how the effects of social media access are conditioned by the policies of even lower-capacity states, our findings broaden the political consequences of governments imposing access frictions at scale. While our experimental results substantiate support-based incentives

[2]The moderating effects of social media align with studies in new democracies showing that other types of mass media affect public opinion by exposing citizens to counter-attitudinal perspectives (Brierley, Kramon and Ofosu 2020; Conroy-Krutz and Moehler 2015; Lawson and McCann 2005; Platas and Raffler 2021).

to restrict access to critical content at the individual level, widespread media restrictions are not always pursued. Prior studies argue that governments may also need the media to mobilize citizens (Gehlbach and Sonin 2014), monitor government officials or opposition leaders (Egorov, Guriev and Sonin 2009; Lorentzen 2014; Qin, Strömberg and Wu 2017), gauge public opinion (Huang, Boranbay-Akan and Huang 2019; Morozov 2011; Qin, Strömberg and Wu 2017), or target repression (Gohdes 2020). We highlight two further factors that could shape decisions to restrict access. We first suggest that partial bans can increase incumbent support by changing the composition of online content production, in line with evidence that censorship alters newspaper and television market equilibria (Kronick and Marshall 2024; Qin, Strömberg and Wu 2018). Conversely, we also find that governments especially lose support among citizens who lost access to news on social media. This finding complements studies showing that citizens are more supportive of governments which permit access to entertaining or 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 Uganda's election-time social media ban to illuminate these subtler considerations, our observational study contributes to a growing literature leveraging rare opportunities to study institutional change in real-time (Callen, Weigel and Yuchtman 2024).

2 Social media in electoral autocracies

Social media is distinct from traditional media in two key respects (Aridor et al. 2024; Zhuravskaya, Petrova and Enikolopov 2020): lower barriers to entry for content producers and the enabling of horizontal interactions between consumers. First, 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 rely on these platforms' 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. Second, the ability to share, comment on, and approve and disapprove content on platforms like Facebook and WhatsApp enables users to spread, endorse, and discuss news and arguments with others.

How might these distinctive features cause social media access to affect support for ruling parties in electoral autocracies? We argue that this depends on the extent to which social media platforms expose users to *alternative news and perspectives*. We first explain why citizens are more likely to encounter content critical of the incumbent via social media than via traditional media in electoral autocracies relative to established democracies. We then move beyond the effect of individual access to social media in isolation (i.e. when content is held fixed) to consider the broader consequences of government policies to restrict access *en masse* for content production and citizen satisfaction—and, in turn, support for ruling incumbents.

2.1 Regime-critical information and perspectives

In democracies with free and independent media, news organizations compete to attract audiences by producing politically-relevant news and commentary, resulting in relatively accurate reporting and a diversity of viewpoints (Mullainathan and Shleifer 2005). While social media often differs in style and presentation, news topics and elite political messaging on these platforms largely overlap with traditional media content (e.g. Harder, Sevenans and Van Aelst 2017; James, Banducci and Cioroianu 2025).[3] The breadth of social media content also sorts users—either consciously or algorithmically—into engaging with like-minded political perspectives (Peterson and Kagalwala 2021; Sunstein 2018) or avoiding politics entirely (Prior 2007). Consequently, social media may do little to expose citizens to novel politically-relevant news or perspectives. Evidence from established democracies indeed finds that social media users are largely exposed to congenial content with limited effects on political attitudes or behaviors (Allcott et al. 2020; Nyhan et al. 2023).

But social media has greater potential to provide novel news and perspectives in electoral autocracies, where ruling parties exert greater control over traditional media sources (Levitsky and Way 2010). Traditional media content tends to favor incumbents because outlets are owned by the ruling party (Enikolopov, Petrova and Zhuravskaya 2011), dependent on state regulation and advertising revenues (Szeidl and Szucs 2021), or sufficiently few to be bought off (McMillan and Zoido 2004). By contrast, elected autocrats face greater costs of identifying and suppressing regime-critical social media content. Whereas rulers can relatively easily identify, target, and control small numbers of critical TV stations or newspapers, few autocratic governments possess the capacity to continuously monitor, remove, or counteract specific social media content or accounts (Roberts 2020)—whether through direct oversight or forcing social media platforms to act as their agents.[4] Social media critics living abroad accentuate these challenges (Nugent and Siegel 2026).

If content critical of the regime is suppressed less on social media than on traditional media (and other information sources people rely on), social media has the potential to expose citizens to less favorable news or commentary about government performance, opposition policy proposals and ideologies, or popular discontent. To the extent that this content is persuasive, whether because information is novel and credible (see Little 2023) or because alternative frameworks for evaluating government performance are compelling (e.g. Schwartzstein and Sunderam 2021), access to social media will cause citizens to update negatively about the government's competence or alignment with citizens' interests. As we illustrate in a simple model in Appendix A.1, such negative updating is likely to reduce support for the ruling party, especially among initial supporters holding more

[3] This relationship is bidirectional, with social media content influencing traditional news organizations (e.g. Hatte, Madinier and Zhuravskaya 2025) and vice versa (e.g. King, Schneer and White 2017).

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

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).

The second distinctive aspect of social media—horizontal engagement between citizens—may amplify these effects of exposure to regime-critical content on social media. In particular, access to social media can facilitate (online and offline) explicit communication (e.g. Barbera and Jackson 2020; Shadmehr and Bernhardt 2017) and generate common knowledge among citizens (e.g. Chwe 2000; Kuran 1989). Both dynamics have the potential to coordinate users' views and vote choices around content encountered on social media (Enríquez et al. 2024).

2.2 Policies to restrict access to social media

The preceding discussion took relatively more critical content on social media in electoral autocracies as given. But, if social media serves as a source of regime-critical content or can help to organize resistance, incumbents are not powerless to counter these challenges to their survival. 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). We next analyze how government policies restricting access to social media may affect public support for the incumbent by shaping content consumption and production.

Ruling parties vary in how they can combat the threat of social media. Roberts (2018) shows how China's high-capacity regime has used tools of fear, friction, and flooding to discourage the production of content that could increase accountability or coordinate collective action and further limit access to, or distract from, content that evades censorship. But few governments possess the requisite capacity for comprehensive monitoring and targeted suppression of sensitive content.

Most elected autocrats rely on blunter tools to restrict consumers' access to social media platforms and periodically sanction producers, rather than suppress or distract from specific content.[5] In extreme cases, this entails banning social media platforms entirely (Miller 2022). Less disruptive censorship strategies increase the costs of accessing social media by imposing taxes (on social media use or mobile data), slowing websites or platforms, or limiting access to censored platforms to citizens with VPNs. Because such frictions can be enforced by a small number of telecommunications companies, over which governments can exert influence, they are easier to administer. Content production laws, including violations of bans, are selectively invoked to (at least temporarily) remove or punish the most threatening and observable producers.

Compared with unrestricted access to social media, partial restrictions could increase support for incumbents through several informational mechanisms. First, restrictions *limit exposure* to

[5]Using data from the *Digital Society Project* (Mechkova et al. 2022), Figure A1 shows that electoral autocracies are, on average, capable of blocking access to specific platforms or shutting off internet access, but actually do so at rates between democratic and closed autocratic regimes.

regime-critical online content. If uncomfortable truths or commentary are persuasive and more prevalent on social media than alternative information sources, 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 could even increase incumbent support if the restrictions prevent persuadable regime supporters from encountering truthful content but still allow regime opponents to become more favorable toward the government when good news is credibly revealed (Heo and Zerbini 2026).

Second, partial restrictions to online freedoms can alter what social media content is produced. In the first place, this could limit critical reporting about the government by reducing social media users' access to online source material (Hatte, Madinier and Zhuravskaya 2025), such as evidence of electoral fraud reported by local or citizen journalists not reported by traditional media. Furthermore, 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, online content is likely to become more favorable toward incumbents. These media market dynamics 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 opposition. In particular, social media may be a valued source of objective news, social interaction, entertainment, consumer purchases, and even business opportunities. Limiting access may then result in *backlash*. For example, 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). Regardless of personal consumption consequences, citizens who value democratic freedoms or draw inferences about the government's type may disapprove of censoring governments (Gläßel and Paula 2020), though others might view this as a signal of the government's strength (Simpser 2013) or be persuaded by government justifications for media restrictions (Conroy-Krutz 2025).

These countervailing effects of government policies to limit social media access—altering what content is produced and consumed as well as eliciting citizen backlash—can be captured by comparing citizens who do and do not lose access to social media. While all citizens may perceive such restrictions as signals of undemocratic tendencies, our model in Appendix A.1 clarifies conditions under which policies that restrict an individual's access to social media affect their support for the ruling party. We thus hypothesize that incumbent support will increase (decrease) among citizens who lose access relative to those who do not when the positive effects of limiting access to regime-critical social media content and inducing more pro-government content production dominate (are dominated by) backlash among individuals losing access to social media.

3 Media and politics in contemporary Uganda

We next provide an overview of electoral politics, media consumption, and social media content and access in Uganda. Figure 1 charts key events and data collection during our study period.

3.1 Electoral context

Uganda has been ruled by Yoweri Museveni and his National Resistance Movement (NRM) party since 1986. Our study focuses on the January 2021 national elections. President Museveni faced his most credible challenge from Robert Kyagulanyi Ssentamu, nicknamed Bobi Wine, a rapper-turned-MP with wide reach through social media platforms. Kyagulanyi represented the newly-founded National Unity Platform (NUP), which rapidly surpassed the Forum for Democratic Change (FDC) as the leading opposition party. Whereas the NRM's traditional electoral base was in rural areas, the NUP built most of its following among young and urban voters (Khisa 2023). Partisan attachment has declined over the past decade, especially among younger voters, with just over half of those aged 18-30 feeling close to a political party in 2022 (Nannozi and Kakumba 2026).

Since the 2021 elections coincided with the COVID-19 pandemic and restrictions on public gatherings, the Electoral Commission mandated that campaigns follow a "scientific" model, using broadcast and online media to appeal to voters, rather than holding mass rallies as usual.[6] However, enforcement was highly uneven: whereas NRM rallies remained common, opposition rallies were violently disbanded.[7] In November 2020, Kyagulanyi was arrested for violating campaigning restrictions, sparking widespread protests whose suppression resulted in 54 reported deaths. In this repressive context, many voters likely anticipated partisan crackdowns and electoral 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 64% of seats, followed by independents (14%), the NUP (11%), and the FDC (6%). The United States described the elections as "neither free nor fair."[8]

3.2 Traditional and social media consumption

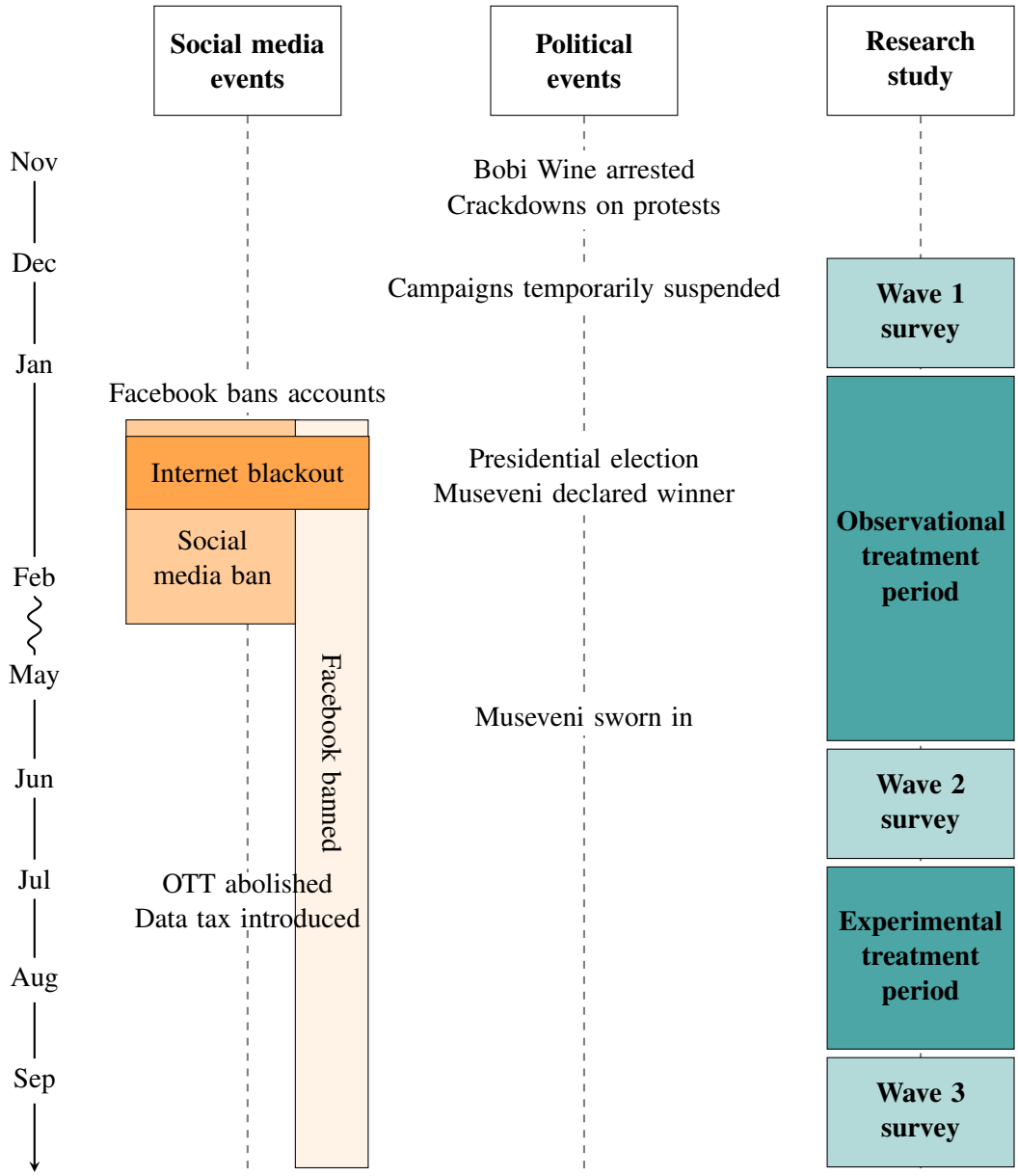
Traditional media remains Uganda's most common source of news. In 2019, 58% of people reported listening to news on the radio at least a few times a week, followed by 34% for television and 12% for print newspapers (Afrobarometer 2019). However, the regime exercises considerable

[6]Uganda Communications Commission, "Market Performance Report Q1 2021"; www.ucc.co.ug/wp-content/uploads/2023/10/UCC-March-2021-Market-Intelligence-Report.pdf.

[7]Freedom House, "Freedom in the World 2021"; freedomhouse.org/country/uganda/freedom-world/2021.

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

Figure 1: Timeline of events during research study



control over such media. 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.[9] Unsurprisingly, our survey respondents perceived radio and TV as notably more supportive than critical of the NRM.

However, social media has become increasingly popular, constituting the vast majority of internet usage. Access is almost entirely through cell phones, with 52% of Ugandans having mobile

[9]US Department of State, "Country Reports on Human Rights Practices: Uganda", 2020; www.state.gov/wp-content/uploads/2021/10/UGANDA-2020-HUMAN-RIGHTS-REPORT.pdf.

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 or political parties. Respondents in our pre-election survey 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: 79% and 78% of our respondents reported using them respectively, while only 17% reported using Twitter.[10]

The political nature of social media content is true across both leading platforms. Among Facebook users in our sample, Figure A2 shows that 71% viewed getting political news as a main reason for using Facebook and 25% cited discussion of current events. For WhatsApp, these figures are 53% and 26%, respectively. 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 (at the time of this study). WhatsApp is also one of the most popular tools for news distribution within communities and campaigning by political parties in the Global South (e.g. Cheeseman et al. 2020; Ventura et al. forthcoming). Political content often originates on more public channels but is shared, discussed, commented on, and reacted to with emojis on both Facebook and WhatsApp. Online 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.

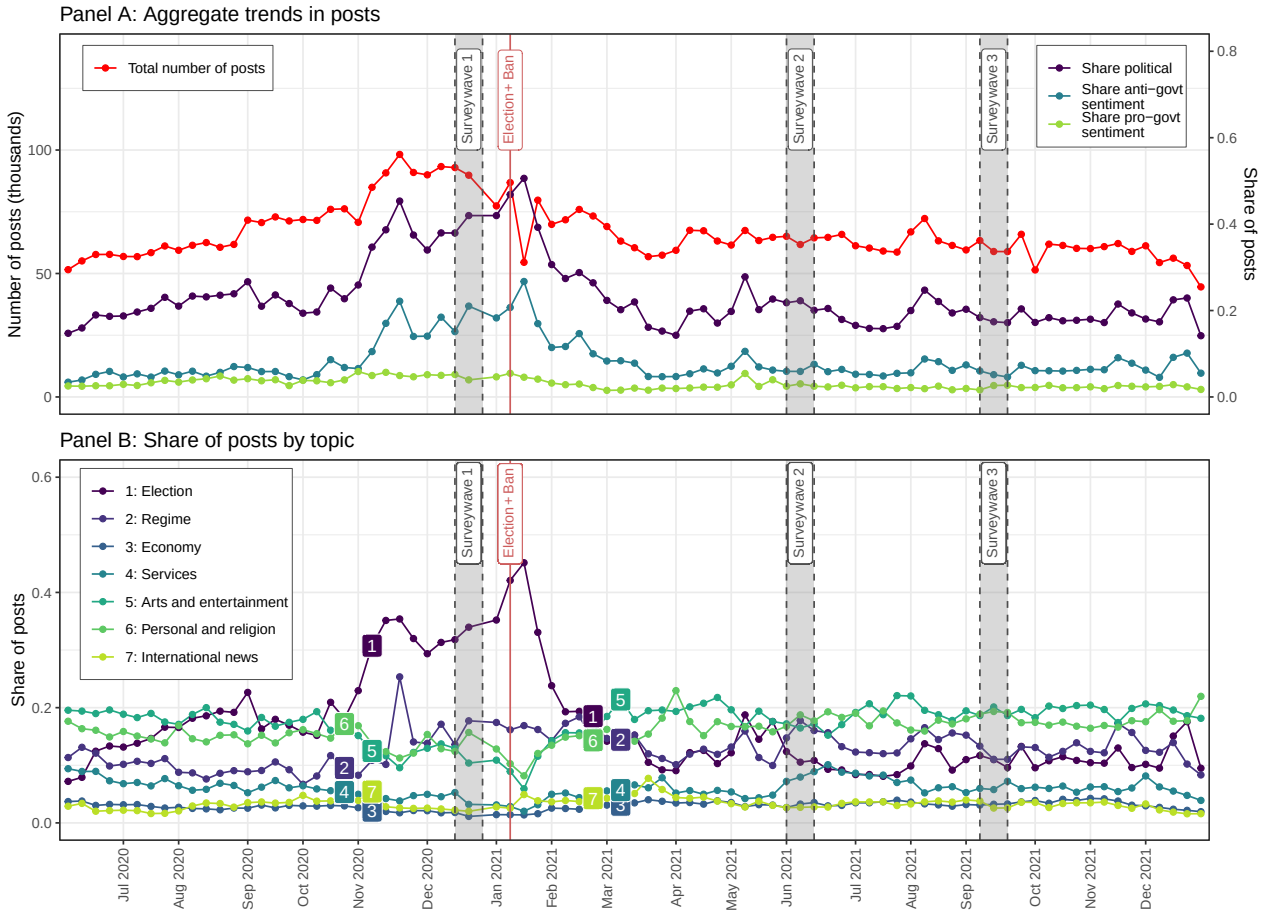
3.3 Social media content during our study

Content on social media in Uganda, especially from the NUP and online news outlets, has generally been critical of the NRM government. To provide a sense of the content our participants were exposed to online during our study period, we focus on Facebook, for which many posts from relatively popular accounts are publicly available and frequently shared beyond the platform. Respondents' similar reasons for using WhatsApp—where content is private and thus unobservable to us—suggest that it played a similar political role in this context (see Figure A2).

We collected a corpus of 6.6 million publicly-available Facebook posts created between June 2020 and December 2021 from the Crowdtangle database. These consist of all posts from a curated set of relatively large political pages and groups pertaining to Ugandan politics (2.6 million posts across 389 pages and groups) and the universe of posts on Facebook pages recorded as administered in Uganda (4.0 million posts across 12,315 pages). For each English language post (83% of the sample), we then classified the presence of Ugandan political content, its sentiment toward the government (anti, neutral, or pro), and its topic using a large language model. Appendix A.2 describes our data and methods in detail.

[10]This is similar to traditional media, with 80% reporting listening to the radio and 72% reporting watching TV.

Figure 2: Facebook posts during study period



Notes: See Appendix A.2 for information on the corpus and its classification. The observational study's treatment period is between survey waves 1 and 2; the experimental intervention occurs between survey waves 2 and 3. See Figure A3 for measures of interactions with posts and Figure A4 for measures using only the corpus of political pages and groups.

Figure 2 shows that social media content was defined by two key features over our study period. First, it was often political in nature. Panel A reveals that 24% of posts were classified as pertaining to politics, peaking at nearly half of all posts before the election. Panel B finds that posts relating to the election quickly dropped from this peak, while posts about the regime (including government institutions, civil rights, protests, the military, and corruption) were common throughout and after the campaign, while posts about the economy, public services, and especially international news were rarer. Second, social media content exhibits a strong anti-government slant. Among posts classified as political, 36% were classified as containing anti-government sentiment and only 13% were classified as pro-government; the rest were classified as neutral. Panel A shows that anti-government sentiment peaked just before the election, reaching a maximum of over half of all political posts—seven times more than the pro-government share of posts.

3.4 Government efforts to restrict access to social media

The government has imposed various policies that limit citizens' access to social media. It first introduced the "over-the-top" (OTT) tax in 2018, which required users to pay 200 shillings ($0.055) per day to access social media platforms, including Facebook, WhatsApp, Instagram, and Twitter. Citizens could evade the tax by using VPNs, but doing so was slower and more data-intensive (Pollicy 2020). 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 the average Ugandan's monthly income.[11] The OTT tax is listed by 55% of our sample as preventing them from using social media more, while 67% cited the cost of data.

Officially motivated as raising revenues and reducing exposure to "gossip" online, these policies are widely seen as frictions intended to undermine opposition support and mobilization (Namasinga and Orgeret 2020)—and indeed reduced social media usage (Boxell and Steinert-Threlkeld 2022). Civil society organizations decried the OTT tax; for example, Amnesty International described it as "a clear attempt to silence dissent, in the guise of raising government revenues."[12]

Around the 2021 elections, the government imposed blunter access restrictions. 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, the Uganda Communications Commission (UCC) announced a ban of all social media platforms, including Facebook, WhatsApp, and Twitter.[13] The ban was lifted on February 10 except for Facebook, which remained officially blocked. On January 13, the UCC shut the internet down completely for five days, resuming only after Museveni was declared the election's winner. Observers described these efforts as attempts to silence opposition voices at a critical moment, undermine efforts to monitor the election, and impede protests.[14]

Many individuals—including government officials—used VPNs to circumvent the ban. Using our Facebook data, Figure 2 documents substantial continued posting once Facebook could only be accessed via VPN, albeit at lower levels than before. We thus characterize the "ban" as introducing significant friction and some degree of fear about the consequences of using social media.

[11] Susi Africa, "Mobile Broadband Pricing Data for Q2 2019. (Source: A4AI Data)", May 30, 2020; susafrica.com/2020/05/30/mobile-broadband-pricing-data-for-q2-2019-source-a4ai-data.

[12] Amnesty International, "Uganda: Scrap social media tax curtailing freedom of expression", July 2, 2018; www.amnesty.org/en/latest/news/2018/07/uganda-scrap-social-media-tax-curtailing-freedom-of-expression.

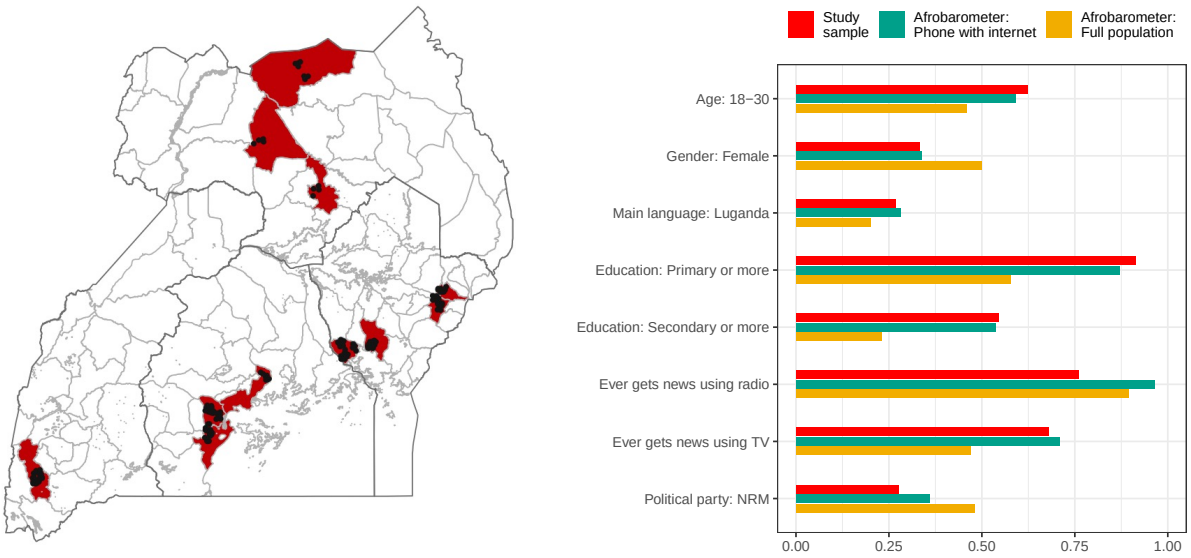
[13] The government implemented the ban by demanding telecommunications companies block access to platforms.

[14] Amnesty International, "Uganda: Authorities must lift social media block amid crackdown ahead of election", January 13, 2021; www.amnesty.org/en/latest/news/2021/01/uganda-authorities-must-lift-social-media-block-amid-crackdown-ahead-of-election. Innocent Anguyo, "Internet and social media shutdowns in Uganda cannot stop growing political resistance", February 3, 2021; blogs.lse.ac.uk/africaatlse/2021/02/03/internet-shutdowns-uganda-museveni-political-resistance-elections-bobi-wine.

Figure 3: Sample characteristics

(a) Sampled districts and trading centers

(b) Demographic covariates



Notes: In Figure 3a, districts are shaded in red and trading centers are black dots. Figure 3b compares our mean baseline survey sample with adult Ugandans from Afrobarometer (2019).

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 social media access could be politically salient and who already used social media, but used 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 3a, 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.

Our research team called 3,710 potential respondents, of which half met our eligibility criteria: aged 18-50; possessing a cell phone able to access social media platforms; and 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 3b 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. The broader population is 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, resurveying 1,310 (85%) wave 1 respondents and adding 145 eligible participants we had been unable to reach for wave 1. Wave 3 was administered in September 2021, 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.[15] Appendix A.5 discusses 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 cared most about the welfare of Ugandans; for our outcomes, we code an indicator for selecting the incumbent NRM party and an indicator for selecting any opposition party.[16] 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) of support for the NRM and opposition parties, which are standardized with respect to control group means and standard deviations.

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

[15] An 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 to instead administer a similar intervention following the wave 2 survey.

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

and non-partisans. Figure 3b shows that this proportion aligns with the share of NRM partisans among Afrobarometer respondents 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 Table A7.

4.3 Social media usage data

To behaviorally measure social media usage, we collected publicly-accessible data on the extent—but not content—of respondents' WhatsApp usage. Although we could not track Facebook activity,[17] we think of WhatsApp usage as a proxy for social media use in general.[18] Figure A2 shows that 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.

We audited the WhatsApp status of the 60% of baseline respondents whose phone numbers were linked to active accounts between four and five times per day throughout the study.[19] 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.[20] Table A7 shows that audited WhatsApp users were 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 reduces citizens' support for the NRM, we conducted a field experiment in mid-2021, following President Museveni's inauguration. In the context of high government-imposed usage costs, our treatment subsidized randomly-selected individuals to use social media. The intervention thus alters individuals' access without meaningfully changing the supply of online content, peers' social media access, or government social media policies.

[17] Tracking Facebook activity was not possible for two reasons: we did not record account URLs for logistical reasons and only a subset of Facebook activity is publicly observable.

[18] During the period of the ban, WhatsApp usage also indicates who was able to circumvent it, and thus able to access social media platforms in general.

[19] 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.

[20] 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.

5.1 Experimental design

After completing the wave 2 survey, we randomly assigned respondents to receive a subsidy to increase their social media use for three months after the elections. Treated participants received payments designed to alleviate 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).

Treated individuals were free to use their mobile data as they liked. Given social media constituted the main use of the internet in Uganda and respondents' stated financial barriers to access, we anticipated the subsidy would principally increase social media usage. The social media content participants were exposed to during the intervention naturally reflects individual and network preferences as well as platform algorithms. As Figure 2 shows, political content on social media was both common and slanted against the government during the intervention period.

To limit differential attrition, individuals assigned to the control condition were instead compensated with a flexible mobile money transfer of 6,000 shillings ($1.69). The average control respondent reported using around half this transfer—which equates to approximately one eighth the value of treatment—to buy data.

Both experimental conditions were 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 Uganda's 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 the following pre-registered OLS regression to estimate average treatment effects in our wave 3 survey data:[21]

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

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 (from 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,

[21] Appendix A.4.3 describes our pre-analysis plan, and explains several deviations with minimal consequences.

Chernozhukov and Hansen (2014), an auxiliary specification adds a LASSO-selected vector of predetermined covariates for robustness. We use robust standard errors for inference.

Since non-NRM partisans are likely to have already been exposed to more regime-critical content than NRM partisans, we pre-specified 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 midline measure of partisanship described above to saturated interactive specifications to estimate differential effects across NRM and non-NRM partisans.[22]

5.2 Effects of the social media subsidy on usage

We first assess the extent to which the intervention affected social media usage, leveraging our high-frequency WhatsApp data as a proxy for social media more generally. To capture treatment effect dynamics, we estimate a panel version of equation (1) in Figure 4.[23] The outcome is the probability that a respondent was "last seen" using WhatsApp on any given day in a week.

The results demonstrate that subsidizing social media consumption 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), from July and August when we sent smaller weekly data transfers (with notifications). The latter generated larger and 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 treatment period, relative to the average probability in 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 increased by 0.09 ($p < 0.01$). Columns (3)-(4) report similar effect magnitudes across NRM supporters and non-NRM supporters, respectively.

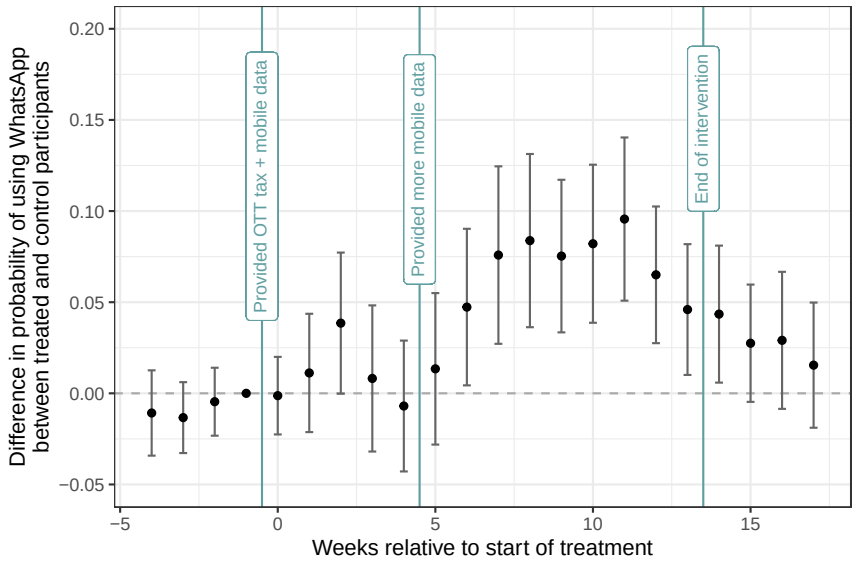
To further characterize effect heterogeneity, we use a causal forest to predict individual-level treatment effects on WhatsApp usage (Wager and Athey 2018).[24] Figure A10 suggests that the subsidy particularly increased usage among somewhat less educated participants, those who reported

[22] For LASSO specifications, we include selected interactive covariates and all their lower-order terms.

[23] These regressions include individual 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 hours spent on social media up to four months earlier.

[24] Our outcome is the difference in the share of days a respondent used WhatsApp during the treatment period relative to before. To predict heterogeneity, we considered predetermined outcomes, the individual indicators comprising them, fixed effects by TC, age, and gender, and respondents' religious, economic, and educational characteristics.

Figure 4: Average treatment effect of the social media subsidy on daily use of WhatsApp



Notes: Estimates are from equation (1), where the outcome is the probability of using WhatsApp on any given day of a week. The baseline category is the week prior to June 1, 2021, when treatment began. The second vertical line marks the beginning of the revised weekly treatment around July 1, 2021, when the mobile data tax replaced the OTT tax. The third vertical line marks the end of treatment around September 1, 2021. All bars indicate 95% confidence intervals. Figure A8 reports day-level estimates.

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 we only audited 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 the social media subsidy 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 indices of party support and their three constituent items.

Our findings for the full sample of respondents document negative but limited average treat-

Table 1: Treatment effects of the social media subsidy on daily WhatsApp usage

	Outcome: Varies by panel			
	(1)	(2)	(3)	(4)
Panel A: Used WhatsApp				
Treatment × Intervention	0.056***	0.054***	0.059***	0.047***
	(0.015)	(0.014)	(0.017)	(0.016)
Treatment × Intervention × NRM supporter			-0.013	0.010
			(0.031)	(0.030)
Treatment × Intervention			0.046*	0.057**
+ Treatment × Intervention × NRM supporter			(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 standard deviation	0.37	0.37	0.37	0.37
Interactive LASSO-selected covariates		✓		✓
Panel B: Number of audited times seen on WhatsApp				
Treatment × Intervention	0.091***	0.083***	0.096***	0.083***
	(0.024)	(0.024)	(0.029)	(0.028)
Treatment × Intervention × NRM supporter			-0.023	0.009
			(0.052)	(0.052)
Treatment × Intervention			0.073*	0.092**
+ Treatment × Intervention × NRM supporter			(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 standard deviation	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. *Treatment* is an indicator for participants' assignment to the subsidy treatment. *Intervention* is an indicator for a given date falling during the subsidy intervention period. 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 captures 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

ment effects of increased exposure to social media content during a period characterized by anti-government slant. Column (1) of Panel A reports a statistically insignificant 0.06 standard deviation reduction in 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

Table 2: Treatment effects of the social media subsidy 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 standard deviation	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 standard deviation	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 standard deviation	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 standard deviation	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 captures the treatment effect among NRM supporting participants. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

about which party 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 documents insignificant

negative effects on respondents being open to vote for the NRM in the future ($p = 0.21$).

However, these average 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 expectation, the conditional average treatment effect captured by the sum of coefficients at the foot of column (3) in Panel A 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 ($p < 0.01$). 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 ($p < 0.05$).

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 reducing frictions to using social media reduced support for the ruling party among initial NRM supporters survives various sensitivity analyses. First, the even-numbered columns in Table 2 demonstrate robustness to adjusting for LASSO-selected covariates. Including these covariates renders the negative average treatment effect marginally statistically significant ($p = 0.10$). Second, Table A10 shows that our estimates are robust to constructing outcome indices 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 TC, suggesting that our results do not reflect interference across participants. Finally, indicating 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 Potential mechanisms

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 persuasive alternative perspectives through social media content. Four additional findings are consistent with this mechanism.

First, in line with this belief updating logic, 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 still-ongoing Facebook ban to access further critical content. Moreover, Figure A9 shows that treatment effects are larger among NRM-supporting respondents predicted—by a causal forest—to have increased their social media use the most in response to treatment.

Second, treatment effects are concentrated among NRM supporters with the greatest prior support for the ruling party and the least prior exposure to opposition perspectives. Table A14 shows that NRM supporters who perceived the central government as performing well before the intervention reduced their support for the ruling party significantly more than those who did not. In line with this being driven by more limited prior exposure to regime-critical news or commentary, the same was true among those NRM supporters who initially reported never having seen information helping them to understand the perspective of opposition parties.

Third, consistent with the persuasive impact of this exposure, respondents—and particularly NRM supporters—reduced their approval of government performance and became more likely to distrust information disseminated by the government (see Figure A7). While these findings align with our interpretation of the results, we do not find significant treatment effects on other components of our prespecified government performance approval index, such as trust in official COVID-19 numbers or agreement with the policy decisions to ban Facebook or tax social media.

Fourth, we do not find evidence that the treatment operated by increasing knowledge of basic political facts. As shown in Figure A7, the subsidy treatment did not systematically lead respondents to learn about current affairs or specific political events that took place during the treatment period, nor to answer more factual questions correctly about politics.[25] Furthermore, Table A14 shows that neither variation in NRM supporters' prior knowledge of recent political events (measured by the share of six 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. That said, measuring a wide range of potentially-relevant political knowledge—which may vary in salience or precision across individuals) during a relatively long treatment period—is challenging, so we cannot rule out that the treatment affected other politically-relevant knowledge not measured in our survey or moderated the effect of treatment in subtler ways.

Together, our experimental results show that increased access to social media reduced ruling party supporters' support for the NRM in the months after Museveni's inauguration, likely due to the persuasive impact of its regime-critical content. This suggests social media could buttress opposition movements in authoritarian regimes by exposing individuals to social media content that

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

is less regime-friendly than government-dominated traditional media markets. At the same time, these results imply that financial frictions to social media access—which existed throughout our study—might bolster political support by depressing usage by regime supporters.

Whether we should expect similar effects of individual social media access during election periods is ambiguous. On the one hand, the intensity of online political content is potentially greater, whereas traditional media faces more acute censorship. On the other hand, at highly salient moments when politics is hard to avoid, marginal access to social media might have a smaller impact on people's exposure to novel perspectives; further, any effects are likely to be conditioned by incumbents' efforts to reshape social media use around election time.

6 Government-imposed social media ban

Our finding that greater social media access reduced prior supporters' favorability towards the ruling party may concern incumbents. Although the Ugandan government likely introduced its month-long social media ban (which continued for Facebook) and five-day internet blackout before the 2021 presidential election to disrupt collective action and election monitoring, the ban provides a rare opportunity to estimate the effects of imposing a major access friction on incumbent support. Such large-scale government policies may have more subtle attitudinal consequences than isolated changes in social media access. While the social media ban limited access to largely-critical content among citizens without VPNs, it could also have disproportionately affected content production by regime opponents afraid of unevenly-applied sanctions and alienated citizens relying on social media for political news, social connection, or economic opportunities. We next explore how the potentially countervailing effects of content shifts and backlash to policy decisions shape the impact of *en masse* government social media restrictions on support for the ruling party.

6.1 Difference-in-differences design

To identify the effects of greater 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 VPN usage. Since non-VPN users could start using VPNs, our design likely underestimates the effect of a fully enforced ban.[26]

[26]Figure A11 reports the baseline distribution of VPN usage. Around 20% of respondents report not using a VPN in the week prior to the wave 1 survey but using one in the week prior to the wave 2 survey. These rates are indistinguishable based on participants' partisanship.

VPN users, unsurprisingly, differ from non-VPN users. Table A18 shows that wave 1 VPN users are younger, use more social media, and favor the opposition. However, VPN users are statistically indistinguishable along most other dimensions, including gender, religion, education, and self-assessed living conditions. Conditioning on location and age, by introducing TC and age fixed effects respectively, largely reduces differences between VPN and non-VPN users to statistical insignificance. To ensure that baseline differences across users are not driving our findings, we exploit within-individual variation over time. Our robustness checks further mitigate against time-varying effects of other differences between VPN and non-VPN users by allowing covariate coefficients to vary across periods and employing various covariate balancing methods.

To estimate the effect of being more likely to retain access to social media during the government's ban, our baseline specification estimates regressions of the following form:

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

where Y_{ict} denotes an outcome for individual i located in TC c in survey period t, $Post\ election_t$ indicates the survey period after social media was blocked by the government, and VPN_i indicates prior VPN users. We include individual fixed effects, μ_i, and period 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 TC level to reflect community-level differences in VPN usage.

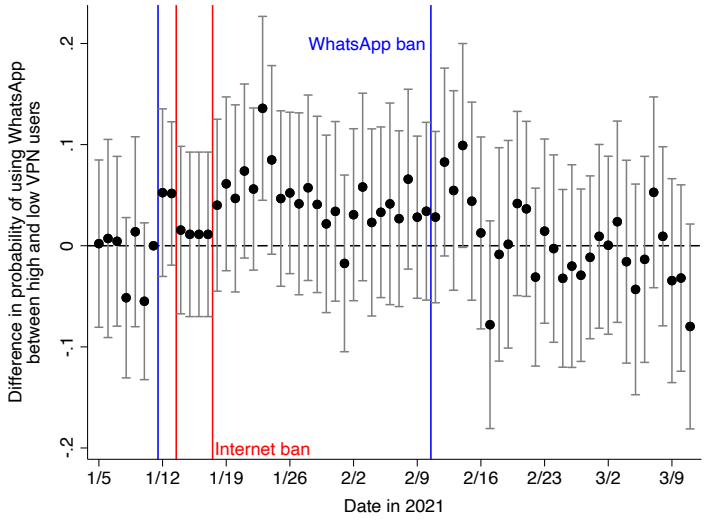
The coefficient τ captures the average differential effect of the election-time social media ban—and its aftermath—on those already using a VPN relative to those who were not under a parallel trends assumption. This assumption requires that VPN and non-VPN users would have followed similar trends in social media usage and government support in the absence of the ban. Figure 5 supports this assumption by showing similar pre-trends in WhatsApp usage across these groups before 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 NRM support *by survey enumeration date* followed similar trends across VPN and non-VPN users over the 18 days of baseline enumeration (concluding seven days before the ban).[27]

6.2 Effects of differential exposure to the ban on social media usage

We first confirm that prior VPN users were more likely to use social media during the ban, using our WhatsApp audit data as a proxy. Figure 5 plots the difference-in-differences estimates by day,

[27]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 5: Differences in daily use of WhatsApp between prior VPN users and non-users



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

relative to the day before the 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 WhatsApp was 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, was thus concentrated during the month after Uganda's presidential election.

Our regression analyses in Table 3 test this relationship by pooling ban and non-ban days. Panel A shows that VPN users were 5.4 percentage points more likely to use WhatsApp on the average day during the social media ban ($p < 0.01$). This relative change in usage is similar in magnitude to our field experiment, although the 31% pre-ban usage rate among non-VPN users exceeded usage in the post-election control group. The post-ban rate among initial non-VPN users indicates that a significant—and, as we find, partisanship-balanced— share of these individuals started to use or used VPNs more regularly, 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.

We again examine heterogeneity using a causal forest.[28] Figure A10 shows that the VPN users who increased their WhatsApp usage most during the ban were relatively better educated, used more social media at baseline, and were less supportive of the government. Table A20 shows that WhatsApp use increased among VPN-using NRM and non-NRM supporters.[29]

[28] We follow the same approach as our experimental analysis except for residualizing our indicator of VPN usage using TC and age fixed effects to obtain a treatment variable more plausibly unconfounded in the cross-section.

[29] In contrast with the field experiment, disaggregating by partisanship poses inferential problems because regime

Table 3: Differential effects of the election-time WhatsApp ban on daily WhatsApp usage by prior VPN use

	Outcome: Varies by panel		
	(1)	(2)	(3)
Panel A: Used WhatsApp			
WhatsApp ban × VPN	0.054***	0.047***	0.046**
	(0.014)	(0.017)	(0.018)
Control mean	0.31	0.30	0.30
Control standard deviation	0.46	0.46	0.46
Panel B: Number of audited times seen on WhatsApp			
WhatsApp ban × VPN	0.084***	0.077**	0.076**
	(0.028)	(0.034)	(0.033)
Control mean	0.54	0.53	0.53
Control standard deviation	0.95	0.95	0.95
Observations	112,943	111,655	111,655
Clusters (TCs)	125	116	116
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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

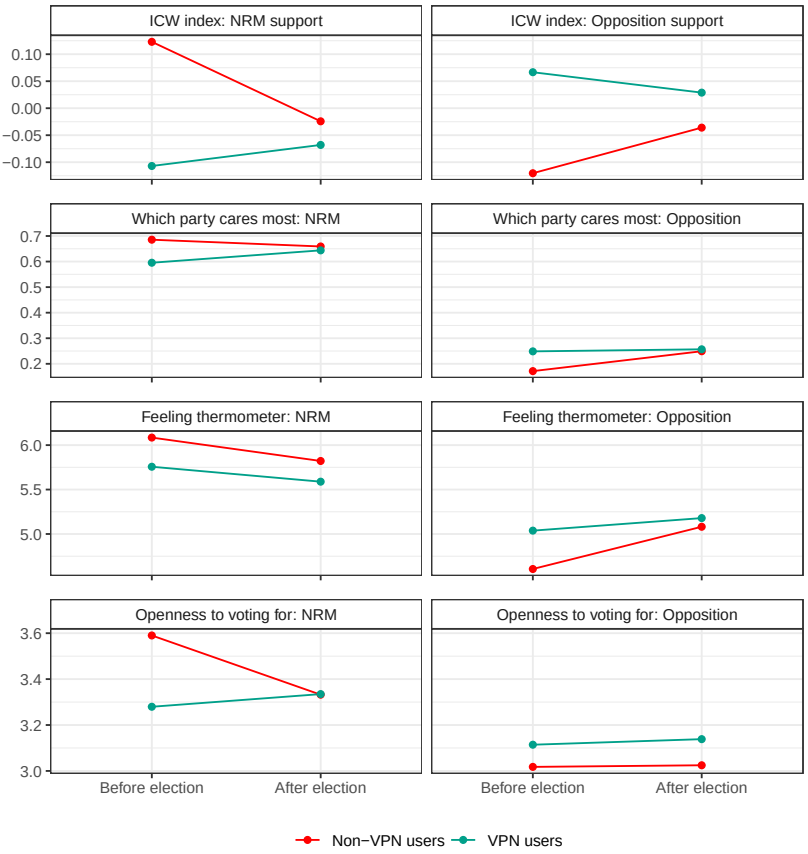
6.3 Effects of differential exposure to the ban on NRM support

We next examine changes in support for the ruling and opposition parties, using the same outcome measures as the field experiment. Figure 6 shows that prior VPN users were initially 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. But 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. Conversely, non-VPN users became more favorable toward opposition parties.

Non-VPN users shifting their support from NRM to opposition parties might suggest that a backlash effect dominates. However, 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 only reveals

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.

Figure 6: Changes in NRM support between pre-election wave 1 and post-election wave 2 surveys, by prior VPN use



Note: Figure plots the raw means of our various outcome measures in the wave 1 and 2 surveys across respondents who did and did not report using a VPN at least one day a week in the wave 1 survey.

that VPN users became more favorable toward the ruling NRM party *relative to non-VPN users*.

We formally test for differential effects of the social media ban by prior VPN use in columns (1) and (4) of Table 4 by estimating equation (2). The results show that the gap between VPN users and non-VPN users significantly shrunk after the ban, with the NRM support index outcome in column (1) of Panel A increasing from wave 1 to wave 2 by 0.18 standard deviations among VPN users relative to non-VPN users ($p < 0.05$). This narrowing is observed across each outcome: VPN users became nearly eight percentage points more likely to believe that the NRM cares most about Ugandans' welfare ($p < 0.05$), slightly 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, the index capturing support for opposition parties in column (4) relatively decreased 0.12 standard deviations ($p < 0.1$). This is driven by reductions in their perceptions of opposition parties caring most about Ugandans' welfare as well as more negative feelings towards opposition parties overall. Acknowledging the econometric difficulties in interpretation (see footnote 29), Figure A13 and Table A21 report somewhat

Table 4: Differential effects of the election-time social media ban on support for the NRM by prior VPN use

	Support for NRM			Support for Opposition		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Support for NRM and opposition party ICW indexes						
Post election × VPN	0.184**	0.198***	0.155**	-0.118*	-0.161**	-0.105
	(0.073)	(0.075)	(0.068)	(0.070)	(0.071)	(0.072)
Observations	2,620	2,612	2,612	2,620	2,612	2,612
Control outcome mean	0.00	-0.00	-0.00	0.00	0.00	0.00
Control outcome standard deviation	1.00	1.00	1.00	1.00	1.00	1.00
Panel B: Which party cares most about people like the respondent						
Post election × VPN	0.075**	0.078**	0.066*	-0.070**	-0.071**	-0.059*
	(0.034)	(0.035)	(0.034)	(0.032)	(0.033)	(0.032)
Observations	2,620	2,612	2,612	2,620	2,612	2,612
Control outcome mean	0.67	0.67	0.67	0.21	0.21	0.21
Control outcome standard deviation	0.47	0.47	0.47	0.41	0.41	0.41
Panel C: Feeling thermometer (0-very cold – 10-very warm)						
Post election × VPN	0.095	0.285	0.209	-0.334*	-0.438**	-0.313*
	(0.196)	(0.208)	(0.218)	(0.172)	(0.181)	(0.181)
Observations	2,620	2,612	2,612	2,620	2,612	2,612
Control outcome mean	5.95	5.95	5.95	4.84	4.85	4.85
Control outcome standard deviation	2.67	2.67	2.67	2.50	2.50	2.50
Panel D: Openness to voting for party (1-not at all – 5-very open)						
Post election × VPN	0.313***	0.258**	0.195**	0.017	-0.044	0.015
	(0.112)	(0.121)	(0.086)	(0.116)	(0.122)	(0.121)
Observations	2,620	2,612	2,612	2,620	2,612	2,612
Control outcome mean	3.46	3.46	3.46	3.02	3.03	3.03
Control outcome standard deviation	1.40	1.40	1.40	1.45	1.45	1.45
Trading center × Post election fixed effects		✓	✓		✓	✓
Covariate × 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

larger effects among non-NRM supporters, although the difference in effects is not statistically significant.

6.4 Robustness checks

In contrast with our small-scale intervention subsidizing access in the months after the election, 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. We next demonstrate the robustness of these estimates to alternative specifications and

interpretations.

We begin by addressing potential parallel trend violations and compound treatment concerns. First, columns (2) and (5) in Table 4 include TC $\times$ period fixed effects to leverage only variation in respondents' VPN use within TCs. These estimates show that our results are not driven by differential trends in NRM support across locations. Second, we address individual-level potential confounding within TCs by adjusting for the interaction between survey wave and ten baseline characteristics—including demographics, news consumption and knowledge, and various partisanship measures. Columns (3) and (6) of Table 4 show similar estimates when including all covariate $\times$ period interactions simultaneously, while Table A22 reports similar results when adjusting for each covariate separately. Third, we also use covariate balancing to account for VPN and non-VPN users differing in these potentially salient ways. Table A23 shows that our results are not sensitive to applying four different weighting schemes designed to minimize differences in pre-ban characteristics. In sum, these tests suggest our estimates are driven by prior VPN use, rather than different trends in NRM support by other characteristics of our sample.

Our findings are also robust across operationalizations of VPN use. Figure A14 reports similar results, across each outcome, when defining VPN users as respondents who used a VPN by any positive number of days in the week before wave 1 enumeration.

An important alternative interpretation is that respondents could have misreported their political views. To account for our difference-in-differences estimates, misreporting would need to vary over time and differentially by prior VPN use. For example, VPN users may fear punishment for violating the ban and inaccurately profess greater support for the NRM following the election to compensate. Alternatively, non-VPN users might be particularly vulnerable to monitoring or intimidation prior to the election, and hence report artificially high levels of support at baseline.

Fortunately, we find little evidence to suggest such measurement biases drive our results. First, we show that the belief that the NRM conducted the survey was rare and does not confound our estimates. Indeed, only 14% of our pre-election survey respondents believed enumerators were sent by the government (13%) or ruling party (1%), with no differences according to baseline VPN usage. Turning to changes in these beliefs after the election, columns (1)-(4) of Table A24 reveal that VPN users became no more likely to believe enumerators had been sent by the government or the NRM. Furthermore, Table A22 shows that our main results hold when adjusting for the interaction between holding such beliefs at baseline and period fixed effects.

Second, our results are not driven by the participants most likely to believe they could be punished for opposing the NRM. To measure participants' perception that the government is monitoring them, we construct an index using responses to survey questions about the extent to which the ruling party monitors voting behavior, how careful they need to be when commenting on politics online, how likely the government would be to identify citizens posting regime-critical content online, and

the likelihood that the government can ascertain citizens' VPN usage.[30] Prior to the ban, VPN users scored actually marginally higher than non-VPN users. More importantly, we find no differential changes in our monitoring index after the election by prior VPN use in Table A24. Furthermore, Table A22 shows our main results are robust to adjusting for the interaction between beliefs about monitoring and period fixed effects, while Table A25 finds similar effects across participants with higher versus lower values of this index or any of its constituent items.

6.5 Potential mechanisms

The effects of access to social media notably differ between the randomized subsidy after Museveni's inauguration and the election-time social media ban. Despite our research designs isolating quantitatively similar changes in social media usage,[31] the former intervention found greater access reduced support for the NRM, whereas the latter found the reverse. Beyond the significant differences in context, we next explore the extent to which changes in social media content and backlash induced by the government's social media ban help to explain these contrasting results.

6.5.1 Differences in social media content

One potential explanation for the different effects is the content users would have encountered on social media. If content was more favorable toward the NRM from the election-time ban until our wave 2 survey than during the post-inauguration experimental period, this could at least partially explain why VPN users became relatively more supportive of the NRM than non-users after the ban was imposed.

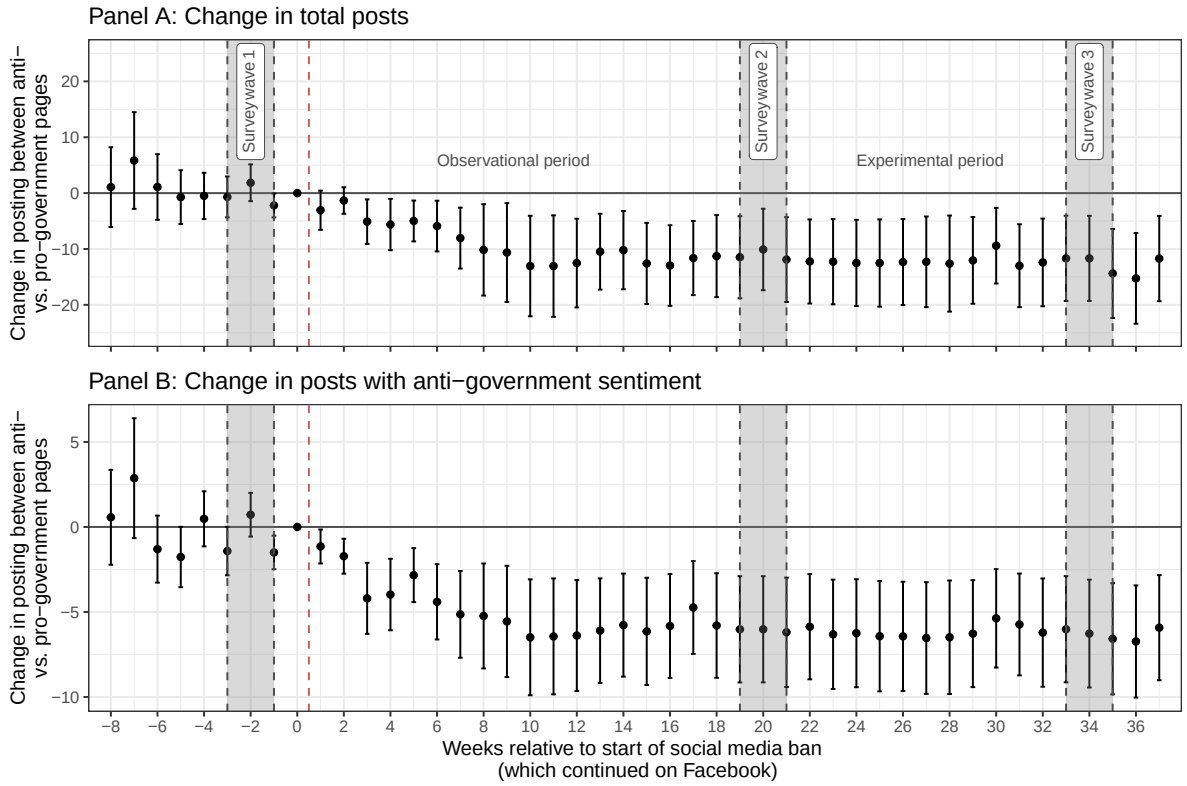
As Figure 2 illustrates, our large corpus of posts from Ugandan Facebook pages and groups were similarly slanted against the NRM, but substantially more political, over the two months following the ban than during the experimental period.[32] But citizens' expectations of anti-government content are likely to at least partly depend on the context. Since violent oppression of opposition protests and leaders in late 2020 and opposition efforts to coordinate online reporting of electoral fraud were common topics of social media posts earlier in the election campaign, citizens would likely have expected extensive reports of violence or fraud that did not ultimately materialize (the reports, that is) during the ban and election's aftermath. The experimental period lacked analogous focal points for revealing the government's type.

[30] The first two of these questions were asked in our wave 1 survey, while the latter two were asked in wave 2. In Table A24 we examine all items asked in each wave as time-varying index outcome; in Table A22 we adjust for an index using items defined at baseline. Table A25 instead reports heterogeneous effects using indexes constructed with either only the items asked at baseline or also including those items asked at midline.

[31] Appendix A.6 suggests the contrasting results are not driven by differences in *whose* social media access changed most in the two designs—whether in terms of sample composition or the characteristics of the effective "compliers."

[32] The day-level Figure A5 shows that aggregate posts mostly recovered quickly after the the ban was imposed.

Figure 7: Changes in posting behavior between anti-government and pro-government pages



Notes: See Appendix A.2 for information on the corpus and its classification. The observational study's treatment period is between survey waves 1 and 2; the experimental intervention occurred between survey waves 2 and 3. 95% confidence intervals plotted.

Although we cannot perfectly distinguish the social media ban's effects from typical electoral cycles in content production, our evidence is consistent with the internet and social media bans rendering social media content less regime-critical than it would otherwise have been. First, the bans appear to have successfully impeded efforts to collect and disseminate categorical evidence of widespread electoral malpractice. As Figure A6 indicates, Facebook posts about electoral fraud, vote rigging, and vote buying barely increased around election day but quickly subsided. This is consistent with the ban cutting off a key basis for social media posts containing regime-critical news or perspectives. Similarly, independent media outlets like the *Daily Monitor* produced few critical articles that could be widely circulated.

Second, even though VPN users were more likely to oppose the government on average, several analyses suggest the social media ban disproportionately chilled the production of regime-critical content. We use a difference-in-differences design to compare changes in social media activity across Facebook pages and groups that we classified as anti- versus pro-government prior to the start of the official election campaign (see Appendix A.2.3 for details). These pages and groups have similar baseline propensities to post about politics (32% and 29% of all posts, respectively).

Table 5: Differential effects of the election-time social media ban on potential mediators of support for the NRM by prior VPN use

	Democracy with major problems		**Follow national government officials**		**Follow opposition politicians**		**Central government performance assessment**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post election × VPN	-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
Control outcome mean	0.55	0.55	0.41	0.41	0.40	0.40	3.28	3.28
Control outcome standard deviation	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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In Figure 7, we examine changes in posting behavior across these pages following the social media ban. Anti- and pro-government pages and groups followed parallel trends during the official election campaign (starting in early November), but posts by anti-government pages—and especially the extent of their regime-critical posts—dropped off significantly and persistently more after the ban. As Table A26 shows, these reductions are robust to exploiting variation only within pages with similar baseline levels of posting activity.

These compositional shifts do not seem to simply reflect anti-government pages and groups uniformly disengaging more after a dispiriting election. To help distinguish these explanations, we leverage our post-level topical coding to show that the reduction in anti-government posting is not purely due to reductions in posting about the election. Indeed, we also find significant reductions in content pertaining to sensitive topics most likely to be sanctioned, with little change in the expression of broader grievances. In particular, Table A27 reports reductions in posting about regime-related topics but no change in discussion of economic conditions or service provision. Disaggregating these regime-related topics, we show in Table A28 that this is driven by fewer posts about security (crime, policing, and repression), corruption, and protests and riots during the period of the observational study. Consistent with the prevention of election monitoring, Table A29 also finds no differential increases in posting about electoral fraud after the ban. These findings, overall, are consistent with the risk of selective enforcement of the ban discouraging critics from producing social media content, and thus shaping a social media environment that was more favorable to the NRM than it was during the election campaign and could otherwise have been afterwards.

Having described some of the changes in the content users likely encountered on social media, we next explore whether users' beliefs shifted in line with this characterization. We start by examining changes in beliefs about the quality of democracy among VPN users, who were more likely to have circumvented the social media ban than non-VPN users, and whether increased support was concentrated among respondents with the initial lowest expectations of fair elections.

Table 6: Heterogeneity in the differential effects of the election-time social media ban on NRM support, by respondent prior beliefs

	NRM support index (ICW)					
	(1)	(2)	(3)	(4)	(5)	(6)
Post election × VPN	0.110	0.125	0.074	0.049	0.071	0.077
	(0.078)	(0.081)	(0.097)	(0.098)	(0.092)	(0.097)
Post election × VPN × Uganda a flawed democracy prior	0.383**	0.391**				
	(0.191)	(0.191)				
Post election × VPN × Followed opposition politicians			0.195	0.272**		
			(0.132)	(0.134)		
Post election × VPN × Non-good incumbent performance prior					0.215*	0.233*
					(0.125)	(0.140)
Post election × VPN + Post election × VPN × 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
Control outcome mean	0.00	-0.00	0.00	-0.00	0.00	-0.00
Control outcome standard deviation	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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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 clearer with covariate adjustment in column (2) ($p = 0.14$). Second, and more tellingly, increased NRM support among prior VPN users after the election is significantly greater among respondents for whom violations of democratic norms were likely to have been less severe than expected. Columns (1) and (2) 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 helps 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.

The marked reductions in anti-government posting during the ban means that individuals retaining access to social media during the ban were exposed to content which was *relatively* more pro-government, and less political in general, than during the pre-election period. Our analyses suggest that VPN users indeed consumed different content and were somewhat swayed by it. Regarding social media content consumption, columns (3)-(6) 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 (7) and (8) 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 are consistent with changes in the slant of online content during the social media ban causing citizens with greater access to social media to update favorably toward the NRM.

6.5.2 Sanctioning of government by citizens harmed by the ban

The preceding analyses suggest that shifts in the nature of online content led VPN users—who were more likely to continue using social media—to view the ruling party somewhat more positively over the course of the social media ban. But the government's policy may *also* have affected the attitudes of those who lost access to social media. While the government's ban restricted everyone's access to social media, its impact on the livelihoods of people not already using VPNs would have been greater. Consequently, the positive difference-in-differences estimate may additionally reflect those who lost access to social media sanctioning the NRM more than those who did not.

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 reduced their ability to find reliable news (58% of respondents answered affirmatively), reduced their ability to talk to friends and family (78%), affected their ability to consume online entertainment content (18%), affected their ability to purchase goods and services (10%), and interfered with their business or job (40%). To investigate the effects of suffering from the social media ban in these ways, we estimate the following difference-in-differences regression:

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

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

The estimates in Table 7 provide evidence of a backlash effect, even several months after social media except Facebook was restored. Columns (1)-(4) show that 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 also 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

Table 7: Differential effects of the election-time social media ban on support for the NRM by censorship experience

	NRM support index (ICW)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Post election × Ban reduced access to reliable news	-0.175***	-0.185**								
	(0.065)	(0.071)								
Post election × Ban reduced social interactions			-0.142	-0.148						
			(0.090)	(0.097)						
Post election × Ban affected access to entertainment					-0.063	-0.054				
					(0.086)	(0.089)				
Post election × Ban affected purchase of goods/services							-0.087	-0.074		
							(0.131)	(0.140)		
Post election × Ban interfered with business/job									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
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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

within-TC variation in column (10). These results thus point to a net sanctioning effect driven by citizens who lost access to reliable news and connections during the ban. Table A30 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 intervention found that several months of greater access to social media during a period characterized by largely regime-critical content reduced support for the ruling NRM party among its prior supporters. However, broad access to such a megaphone can motivate autocratic governments to impose significant frictions on citizens' access. Leveraging a difference-in-differences design around Uganda's election-time social media ban to illuminate the consequences of such policies for ruling party support, we find that a government policy which restricted social media access *en masse* likely reduced the production of social media content that was critical of the ruling party, in addition to reducing exposure to it, and induced backlash among citizens who lost access.

By providing rare causal evidence from an electoral autocracy, these results point to distinct consequences of social media access for political attitudes outside established democracies. First, social media has the potential to serve as a democratizing force by reducing support for elected autocrats, in contrast with findings from established democracies, where social media often informs users without affecting political views on average. These counter-attitudinal effects are consistent with traditional media content in electoral autocracies exposing citizens to less critical information,

alternative opinions, or opposition messaging than in democracies, as exemplified by the prevalence of opposition posts on social media. This is likely because elected autocrats find it harder to control the production of regime-critical content on social media, whereas low barriers to entry that permit diverse content are common across traditional and social media in established democracies. Social media in electoral autocracies is thus politically important not simply because it exposes citizens to more critical news and perspectives than traditional media do, but because many incumbents cannot fully control it.

Second, the consequences of social media access for citizens' political attitudes are conditioned by the policies of electoral autocracies with limited capacity to surveil, remove, or out-compete sensitive content. Extensive restrictions have the potential to redefine the composition of social media content, but may also upset regular users. By contrast, using fear and friction to shape who engages with what social media content is constrained by stronger laws and norms in established democracies; political actors are largely left with flooding, although competition for attention may limit its effectiveness. Ultimately, restricting access to social media can therefore create trade-offs between controlling political discourse and maintaining popularity for elected autocrats.

The implications for social media strategies in low-capacity electoral autocracies are likely to depend on how incumbents balance preventing collective action (e.g. Roberts 2018) and maintaining popularity (e.g. Guriev and Treisman 2020). Elected autocrats principally concerned about collective action may tolerate some of the backlash we document in order to restrict social media access that could coordinate mass action capable of catalyzing popular overthrow or elite coups. But, as our Ugandan case suggests, the political costs may explain why heavy social media restrictions are only deployed at key moments of potential mobilization, such as elections or protests. For leaders primarily concerned about popularity, the trade-off is more subtle. Here, sustained fear and frictions are required to maintain incumbent support by continually limiting the production of regime-critical content and reducing exposure to it. However, lasting and widespread restrictions are also likely to reduce support, by both revealing the incumbent's authoritarian tendencies and reducing the welfare of citizens who value access to news and communication (Kronick and Marshall 2024). This may be why some elected autocrats possessing only blunt instruments of censorship impose lower frictions on social media access than analysts might expect.

However, the backlash constraint may be alleviated for autocratic regimes with greater capacity to use targeted digital tools. Regimes like China, and to a lesser extent Russia, can identify and remove specific pieces of content or accounts and prevent workarounds in ways that are less visible and disruptive to ordinary users, potentially limiting the widespread backlash that accompanies blunter restrictions. Uganda's lower capacity is reflected in the incompleteness of its social media ban and the need for internet blackouts at sensitive moments. Yet, as technology develops and governments like China export social media censorship tools to allied regimes, the costs of more

sophisticated strategies mixing fear, friction, and flooding may diminish (Roberts 2020). This may ultimately undercut social media's potential to reduce support for elected autocrats by increasing the subtlety of censorship and the quality of distracting content.

Our analysis of the Ugandan case—including a rare opportunity to assess the political consequences of repressive social media policies—provides a stark setting that helps to clarify some of the key trade-offs governing an autocrat's censorship decisions. But how far do the results generalize? Within Uganda, our results pertain to a sample of younger, peri-urban, and occasional social media users in competitive districts. Our results suggest that more nascent users could experience larger drops in ruling party support, to the extent that they are otherwise more likely to support the ruling party and less exposed to regime-critical content. This could explain why modest frictions to social media access have remained in place since 2018. Beyond Uganda, other electoral autocracies like Russia, Tanzania, or Venezuela also do more to restrict traditional than social media. But the threat of social media in Uganda may also be unusual, since critical commentary and opposition politics are especially youth-led and social-media-centered.

Future research might usefully expand on the conditions under which social media weakens elected autocrats, when governments can neutralize that threat, and when restrictions backfire. First, since we cannot establish exactly what content individuals consumed on social media, a key question is what types of content persuade consumers—and how prevalent they are for different types of users. Parsing whether users are, for example, influenced by news about government performance, alternative worldviews and values, novel narratives about events, or horizontal persuasion by peers is likely to require research designs controlling specific content exposure, rather than our approach of capturing overall effects of social media access when respondents are free to consume what they want. Second, future work could distinguish effects across platforms and user populations. Different platforms facilitate different combinations of original content creation, news sharing, and political deliberation, while the effects we estimate among marginal users may not generalize to older generations of new users who might experience social media differently. Third, more research is required to evaluate how citizens sanction and adapt to the more targeted censorship or flooding strategies that are likely to emerge as technological capacity improves. Informational experiments might help to distinguish the effects of the policy from content encountered on social media, and thus gauge the magnitude and durability of the backlash elected autocrats face when deploying different modes of digital repression.

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):629–676.

Allcott, Hunt, Matthew Gentzkow, Winter Mason, Arjun Wilkins 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.

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.

Arceneaux, Kevin, Martial Foucault, Kalli Giannelos, Jonathan Ladd and Can Zengin. 2024. "Facebook increases political knowledge, reduces well-being and informational treatments do little to help." *Royal Society Open Science* 11(10):240280.

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. 2024. "The economics of social media." *Journal of Economic Literature* 62(4):1422–1474.

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.

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.

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." National Bureau of Economic Research working paper 26567.

Callen, Mike, Jonathan L. Weigel and Noam Yuchtman. 2024. "Experiments about institutions." *Annual Review of Economics* 16:105–131.

Castells, Manuel. 2015. *Networks of Outrage and Hope: Social Movements in the Internet Age*. Polity Press.

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.

Cheeseman, Nic, Jonathan Fisher, Idayat Hassan and Jamie Hitchen. 2020. "Social media disruption: Nigeria's WhatsApp politics." *Journal of Democracy* 31(3):145–159.

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. 2025. "Muzzling the Media? Explaining Popular Support for Media Restrictions in Africa." *Journal of Politics* 87(3):1169–1181.

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.

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.

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.

Gauthier, Germain, Roland Hodler, Philine Widmer and Ekaterina Zhuravskaya. 2026. "The Political Effects of X's Recommender Algorithm." *Nature* (652):416–423.

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.

Gohdes, Anita R. 2020. "Repression technology: Internet accessibility and state violence." *American Journal of Political Science* 64(3):488–503.

Guess, Andrew M., Neil Malhotra, Jennifer Pan 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.

Harder, Raymond A., Julie Sevenans and Peter Van Aelst. 2017. "Intermedia Agenda Setting in the Social Media Age: How Traditional Players Dominate the News Agenda in Election Times." *International Journal of Press/Politics* 2:275–293.

Hatte, Sophie, Etienne Madinier and Ekaterina Zhuravskaya. 2025. "Conflict Reporting in the Digital Age." Working paper.

Heo, Kun and Antoine Zerbini. 2026. "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.

James, Charlotte, Susan Banducci and Iulia Cioroianu. 2025. Flows of Information in Election Campaigns: Who Influences Whom? In *Media and the British General Elections of 2015-2019*, ed. Susan Banducci, Laszlo Horvath, Ekaterina Kolpinskaya and Daniel Stevens. Edinburgh University Press pp. 60–82.

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

Khisa, Moses. 2023. "Uganda's ruling coalition and the 2021 elections: Change, continuity and contestation." *Journal of Eastern African Studies* 17(1-2):325–343.

King, Gary, Benjamin Schneer and Ariel White. 2017. "How the news media activate public expression and influence national agendas." *Science* 358(6364):776–780.

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).

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.

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, Tiago Ventura, Shelley Liu, Carolina Torreblanca and Joshua A. Tucker. 2026. "Reducing Social Media Usage During Elections: Evidence from a Multi-Country WhatsApp Experiment." Working paper.

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.

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

Meserve, Stephen A. and Daniel Pemstein. 2018. "Google politics: The political determinants of Internet censorship in democracies." *Political Science Research and Methods* 6(2):245–263.

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. 2011. *The Net Delusion: The Dark Side of Internet Freedom*. PublicAffairs.

Mullainathan, Sendhil and Andrei Shleifer. 2005. "The market for news." *American Economic Review* 95(4):1031–1053.

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.

Nannozi, Mabel and Makanga Ronald Kakumba. 2026. "Indifferent or disempowered? Ugandan youth lag behind older citizens in political and civic engagement." *Afrobarometer Dispatch No. 1165* .

Nugent, Elizabeth R and Alexandra A Siegel. 2026. "How Exiles Mobilize Domestic Dissent." *The Journal of Politics* 88(2):647–665.

Nyhan, Brendan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak 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.

Qin, Bei, David Strömberg and Yanhui Wu. 2024. "Social media and collective action in China." *Econometrica* 92(6):1993–2026.

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.

Schwartzstein, Joshua and Adi Sunderam. 2021. "Using models to persuade." *American Economic Review* 111(1):276–323.

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 E. Roberts and Pablo Barberá. 2017. "From liberation to turmoil: social media and democracy." *Journal of Democracy* 28(4):46–59.

Ventura, Tiago, Rajeshwari Majumdar, Jonathan Nagler and Joshua A. Tucker. forthcoming. "Misinformation Exposure Beyond Traditional Feeds: Evidence from a WhatsApp Deactivation Experiment in Brazil." *Journal of Politics* .

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.

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

Contents

A.1 Formalizing effects of access to social media A1
A.1.1 Effects of access to social media A1
A.1.2 Effects of introducing a social media ban A2
A.2 Ugandan Facebook data A3
A.2.1 Corpora of Facebook posts A3
A.2.2 Classifying Facebook posts A4
A.2.3 Classifying Facebook pages A8
A.3 Constructing our panel survey sampling frame A8
A.4 Field experiment design A9
A.4.1 Randomization and treatment conditions A9
A.4.2 Automated selection of covariates A9
A.4.3 Deviations from pre-analysis plan A9
A.5 Research ethics A11
A.6 Reconciling results based on complier characteristics A12
A.7 Additional Figures A14
A.8 Additional Tables A29

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 by assuming 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 (which adjusts for context, e.g. people may expect to hear about electoral malfeasance around elections), 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 consuming social media τ_i is decreasing in prior support μ_i for the incumbent. These predictions generate our hypotheses regarding the average and heterogeneous effects of subsidizing access to social media.

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

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, our model of 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 change in support (e.g. because the ban revealed 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:

$$\begin{aligned} \mathbb{E}[\tau_{DiD}] &:= (\mathbb{E}[\tilde{\mu}_i | i = V] - \mathbb{E}[\mu_i | i = V]) - (\mathbb{E}[\tilde{\mu}_i | i = N] - \mathbb{E}[\mu_i | i = N]) \quad (4) \\ &= [(\alpha_V - \alpha_N)m + (\alpha_V \mu_V - \alpha_N \mu_N)] + \beta_V(s + \Delta_s - \mu_V) + \Delta_c. \quad (5) \end{aligned}$$

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 isolation, 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 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.

[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.

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 4.0 million posts throughout the study period from 12,521 distinct Facebook pages.

A.2.2 Classifying Facebook posts

With this aggregate corpus of 6.6 million Facebook posts, we executed three classification tasks to define (A) whether a post is *political* or not (i.e., pertaining to politics in Uganda), (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).

To accomplish these tasks, we used a Large Language Model (LLM) to classify the full set of posts. Balancing classification accuracy against cost given the size of the corpus, we used OpenAI's

recent `gpt-5-nano` model. Consistent with a number of recent contributions, we find the GPT classification performance to perform extremely well (Gilardi, Alizadeh and Kubli 2023; Ziems et al. 2023).

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

Partisanship	MP Candidates	Media	Group	Page
NRM	0.02	0.06	0.08	0.04
NUP	0.28	0.24	0.28	0.44
Other opposition	0.20	0.06	0.22	0.17
Independent	0.09			
Government				0.01

Table A4: Comparison of classifications across corpora

Corpus	Political	Anti-government	Neutral	Pro-government
1	0.50	0.20	0.73	0.07
2	0.11	0.03	0.96	0.01

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. For example, while 2% of the posts made by NRM candidates contained anti-government sentiment, 28% of those made by NUP candidates did. Table A4 intuitively documents a much higher share of *political* posts in Corpus 1 (50%) than in Corpus 2 (11%).

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 LLM-based classification 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 tasks, we find our classifications to perform extremely well.

In panel A, we find accuracy measures of between 0.87 and 0.89 between our GPT 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 similarly high accuracy scores on the anti-government sentiment classification task of between 0.84 and 0.87, again with slightly greater accuracy in the less-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

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							
LLM	Coder 1	0.87	0.91	0.84	0.87	0.91	0.95
LLM	Coder 2	0.89	0.92	0.86	0.88	0.91	0.95
Coder 1	Coder 2	0.91	0.94	0.87	0.90	0.94	0.97
B. Classifying posts as anti-government							
LLM	Coder 1	0.84	0.90	0.81	0.86	0.88	0.93
LLM	Coder 2	0.87	0.91	0.84	0.89	0.89	0.94
Coder 1	Coder 2	0.89	0.93	0.86	0.90	0.92	0.95

classification and 0.89 for the more complex task of classifying sentiment. These benchmarks suggest that our LLM classifier was around 3% less accurate than a hand-coded benchmark, while covering a vastly greater scale of data.

For task (C), to discern the topic of each post, 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 had the LLM 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. government, government ministries, government policies;
3. economy, wages, labor, inflation, currency, poverty, unemployment;
4. crime, security, justice, prisons, police, military;
5. corruption, bribery, embezzlement;
6. protest, unrest, riot;
7. voting, elections, rallies;
8. election fraud, rigging, vote buying;
9. education, exams, schools, universities;
10. health, public health, healthy living, health care;
11. infrastructure, roads, transport, cars, traffic, traffic accidents;
12. water, sanitation;
13. electricity, power;
14. farming, agriculture, crops;
15. social programs, policies to reduce poverty;
16. science, technology;
17. land, expropriation, property rights;
18. human rights, journalism, freedom of speech, civil liberty;

19. weather, environment, extreme weather, natural disasters, climate change;
20. arts, culture, society, music, festivals, food, fashion, lifestyle;
21. entertainment, celebrities, showbiz;
22. careers, individual business activities;
23. sports, athletics;
24. advertisements;
25. personal stories, life updates;
26. religion, church, worship, gospel;
27. 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 A6 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	0.98	0.38	0.99	0.42	0.94	0.26
Regime	0.84	0.36	0.92	0.47	0.72	0.19
Economy	0.22	0.04	0.44	0.11	0.13	0.02
Services	0.28	0.03	0.43	0.06	0.21	0.02
Arts and entertainment	0.01	0.00	0.02	0.01	0.00	0.00
Personal and religion	0.01	0.00	0.04	0.01	0.00	0.00
International news	0.07	0.02	0.19	0.06	0.03	0.00
Other	0.00	0.00	0.00	0.00	0.00	0.00

A.2.3 Classifying Facebook pages

Our analysis of posting behavior across pages (and groups) before and after the social media ban (see Section 6.5.1) requires a categorization of pages/groups as being anti-government or pro-government in both corpora. To do this, we subset to the period prior to the start of the election campaign (though results are invariant to including the campaign period as well) and define a page/group as being *anti-government* if: (1) A non-zero proportion of its posts are classified as political; (2) A greater share of its posts are classified as being anti-government than pro-government. We define a page/group as being *pro-government* if condition (1) holds but condition (2) does not and *neutral* when neither condition holds.

This classification leads 76% of pages/groups to be defined as *neutral*, 13% as *anti-government*, and 11% as *pro-government*. As noted in the main text, using this classification, both anti-government and pro-government pages/groups post roughly equal proportions of political content at baseline (32% and 29%, respectively) although anti-government pages/groups are substantially more active, with an average of 401 total weekly posts compared to 225 among pro-government pages/groups. Our analyses exclude the neutral pages/groups since they represent a less appropriate comparison group while also posting far less often than the other types.

Validating this using our hand-coded classifications of partisanship in Corpus 1, reassuringly only 4% of NRM-coded pages/groups are classified as being anti-government while only 5% of NUP-coded pages/groups are classified as being pro-government.

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 or \$1.69); and (ii) provided 1.5GB of mobile data for the month for Airtel users (UGX 10,000 or \$2.82) or 500MB a week for MTN users (UGX 5,000 or \$1.41 a week for four weeks). In July and August, treated Airtel users received 400MB a week (UGX 3,500 or \$0.98 a week) and treated MTN users received 500MB a week (still costing UGX 5,000 or \$1.41 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 (\$1.69) 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 A7c 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 A7b 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.

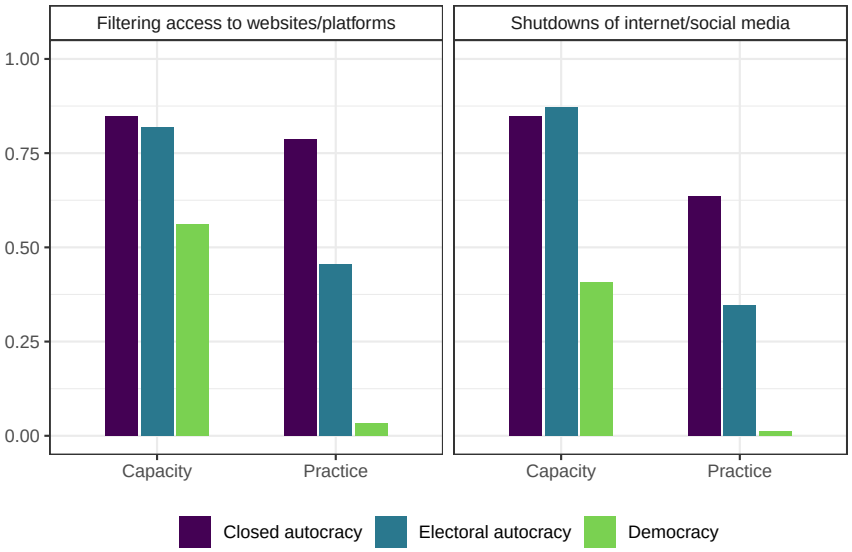
Second, we consider subtler differences in the characteristics of the effective "compliers" across the two designs—i.e., whether different types of participant were induced to increase their social media usage across the designs—which might then explain differences in their subsequent political

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.

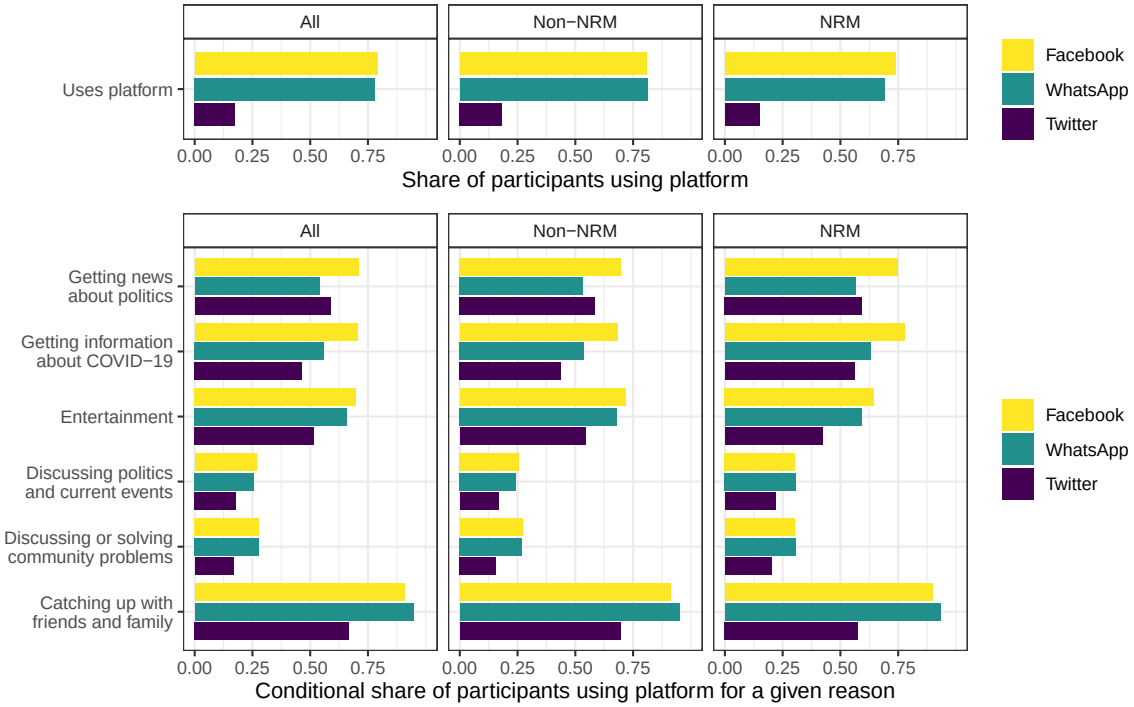
A.7 Additional Figures

Figure A1: Regime types and decisions to restrict online access



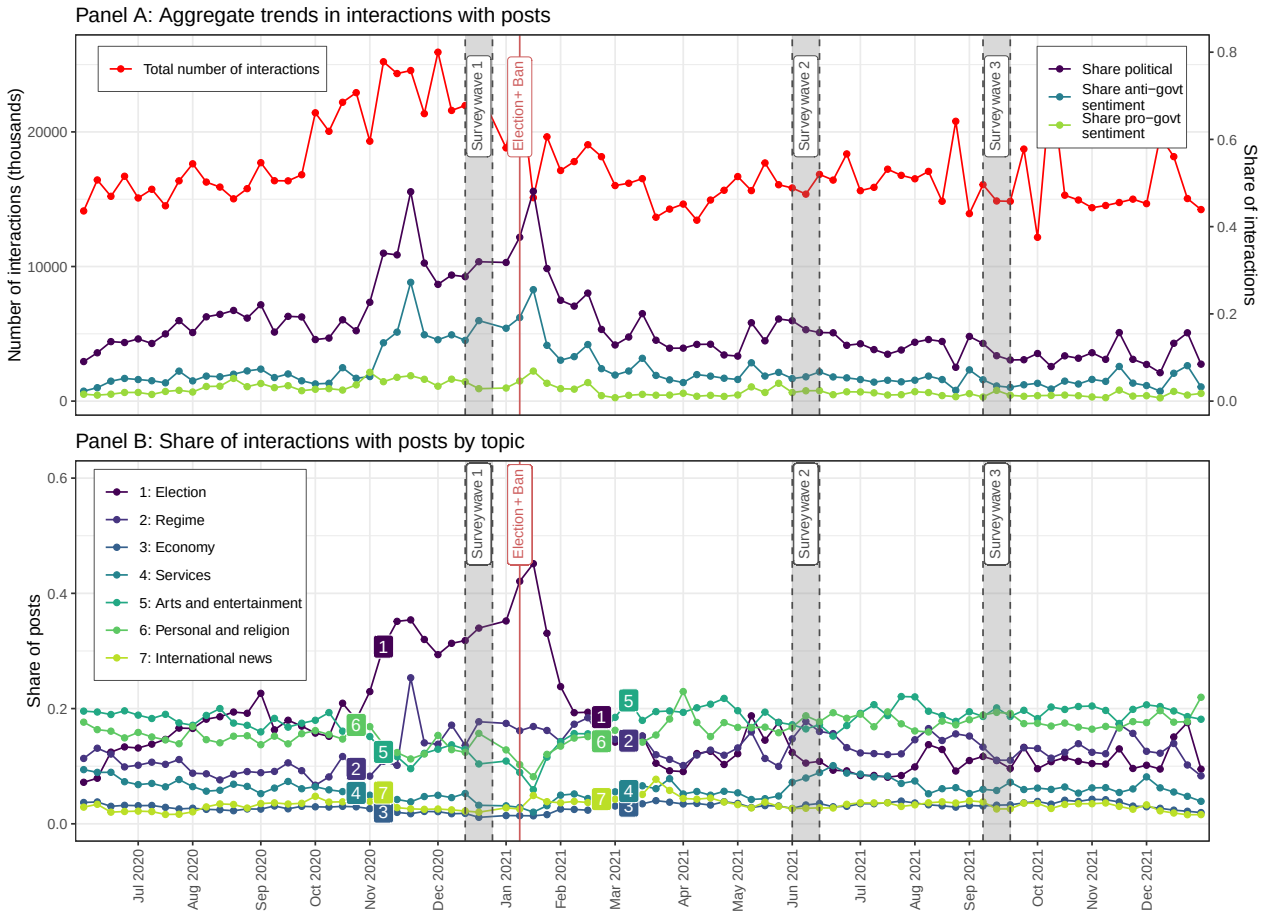
Notes: Figure uses data from 2023 edition of *Digital Society Project* (see Mechkova et al. 2022). V-Dem regime classifications used, aside from pooling *Electoral democracies* and *Liberal democracies*. Filtering capacity: Government has technical capacity to censor information by blocking access to 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



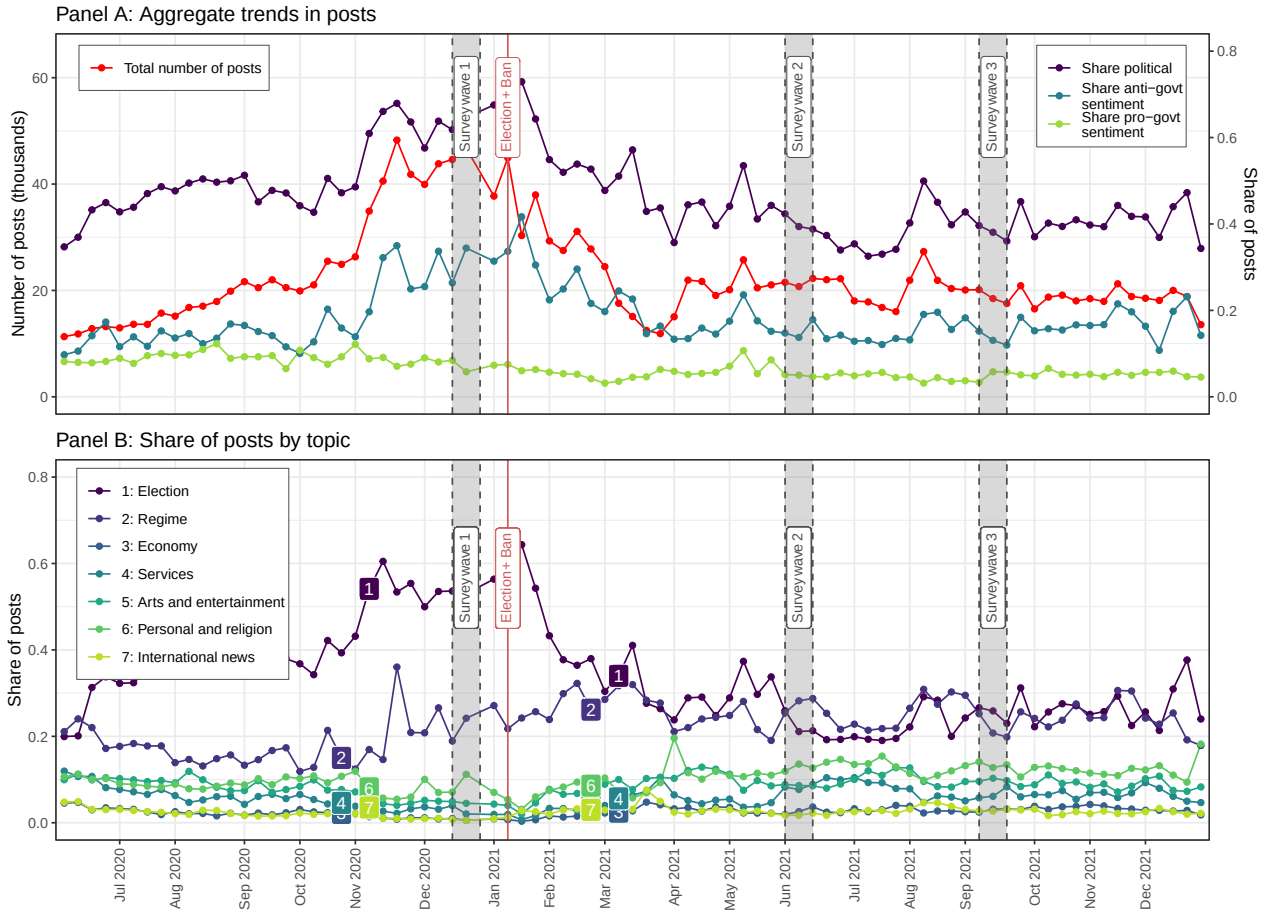
Notes: The top panel reports participants' self-reported use of different social media platforms, overall and by partisanship. The bottom panel reports the main reasons for using a platform conditional on those who report using a given platform.

Figure A3: Facebook interactions data during study period



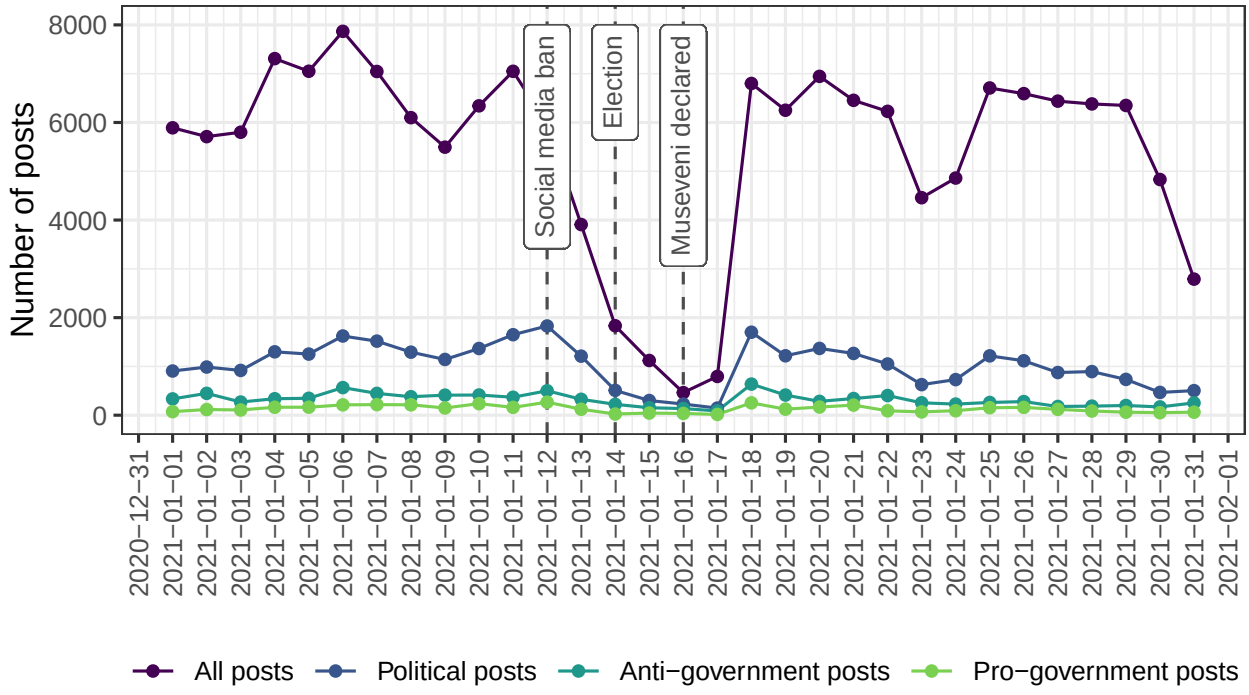
Note: See Appendix A.2.1 for information on corpus of posts and Appendix A.2.2 for information on coding. Crowdtangle-defined 'interactions' comprise likes, reactions, comments and shares.

Figure A4: Facebook post data during study period (Corpus 1)



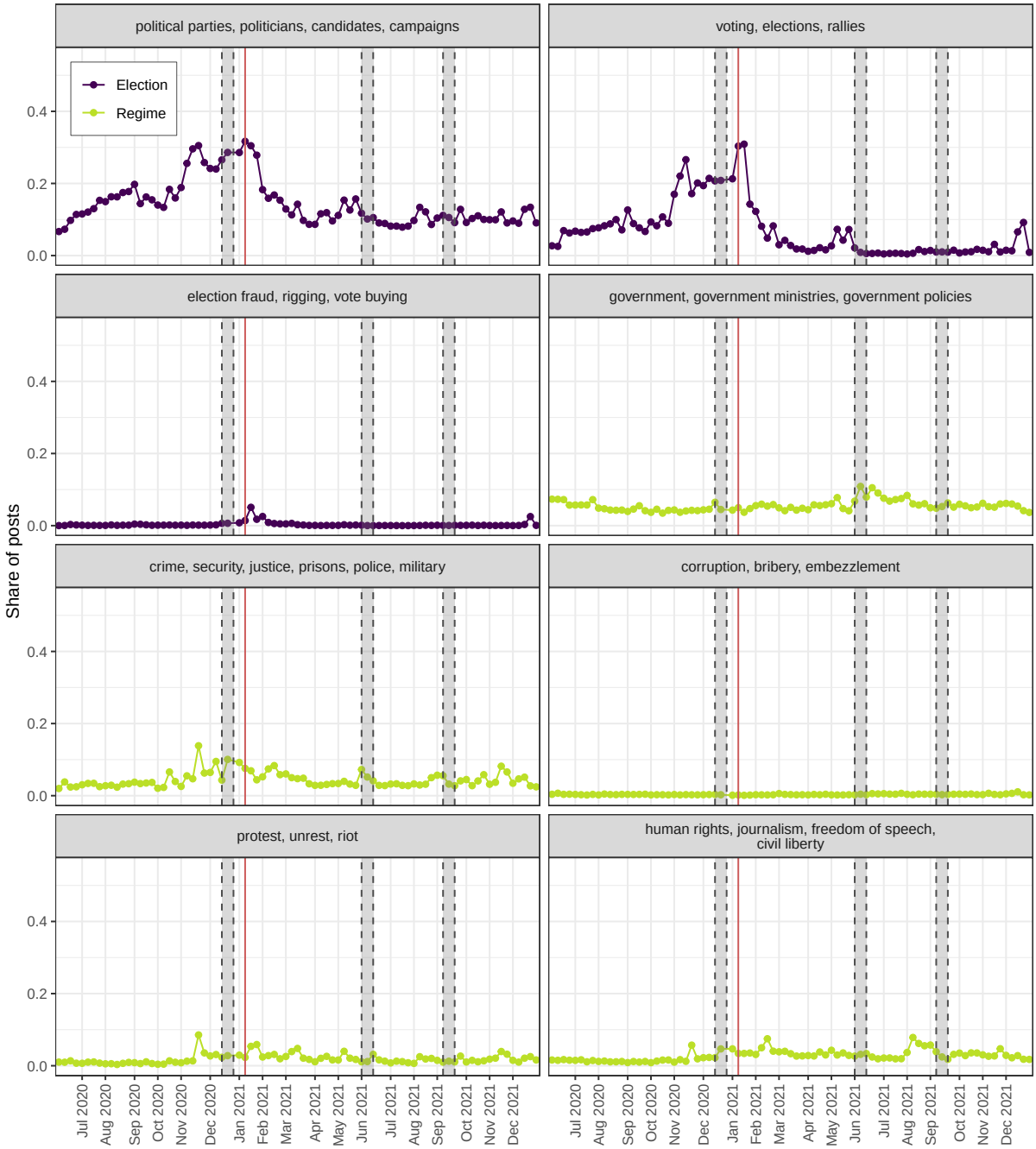
Note: See Appendix A.2.1 for information on corpus of posts and Appendix A.2.2 for information on coding.

Figure A5: Day-level Facebook posts around election day

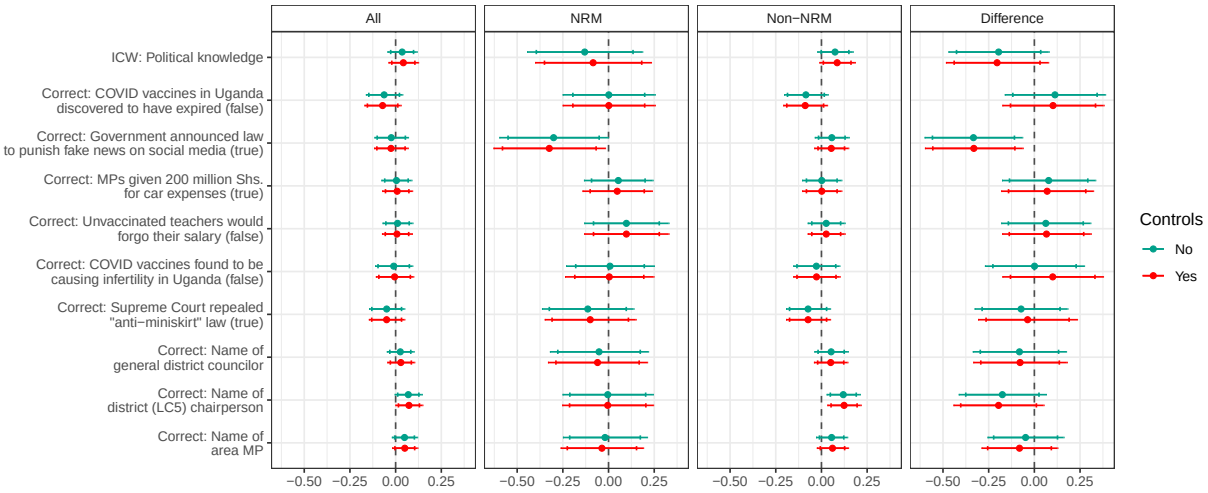


Notes: See Appendix A.2.1 for information on corpus of posts and Appendix A.2.2 for information on coding. The social media ban was introduced on January 12, the internet blackout began on January 13, the election was held on January 14, and the internet blackout was lifted on January 18.

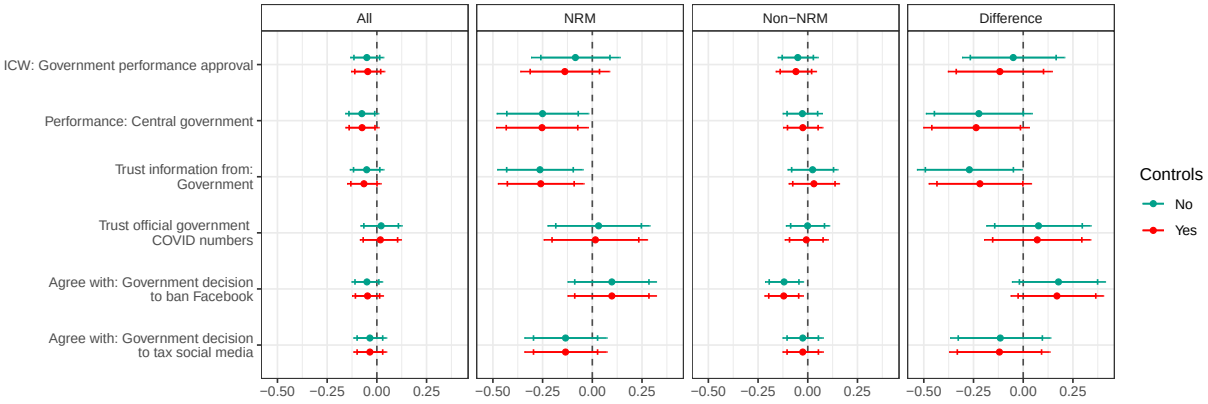
Figure A6: Disaggregation of Facebook post topics



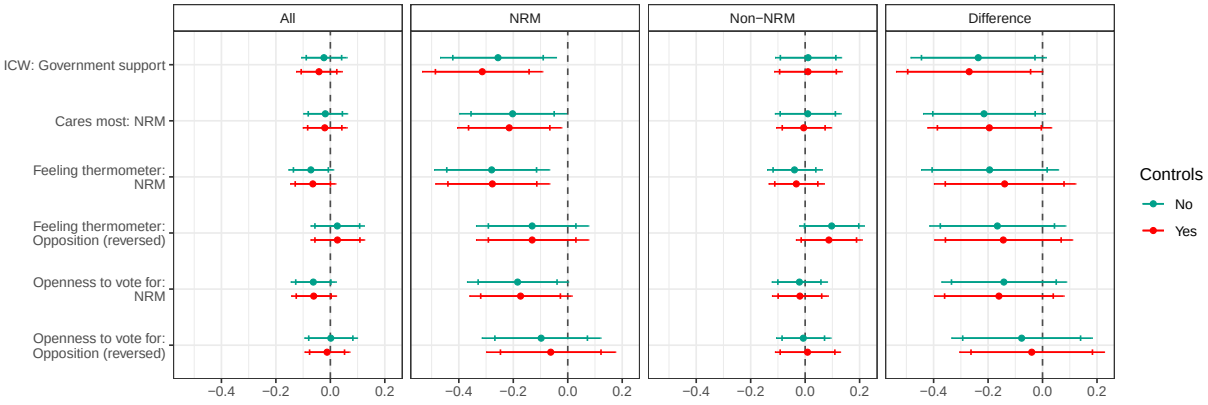
Note: See Appendix A.2.1 for information on corpus of posts and Appendix A.2.2 for information on coding.



(a) Knowledge of current political events



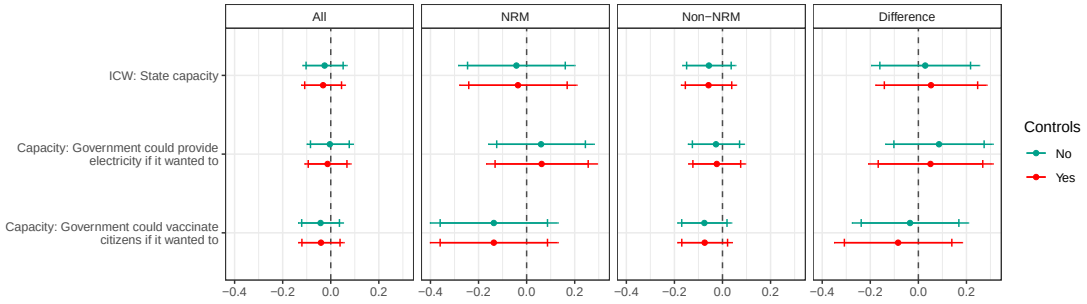
(b) Government performance approval



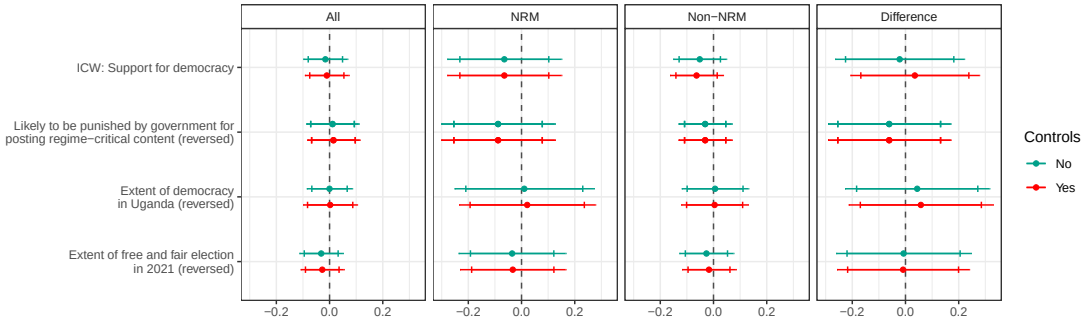
(c) Government support

Figure A7: Treatment effects of experimental social media subsidy on pre-registered indices

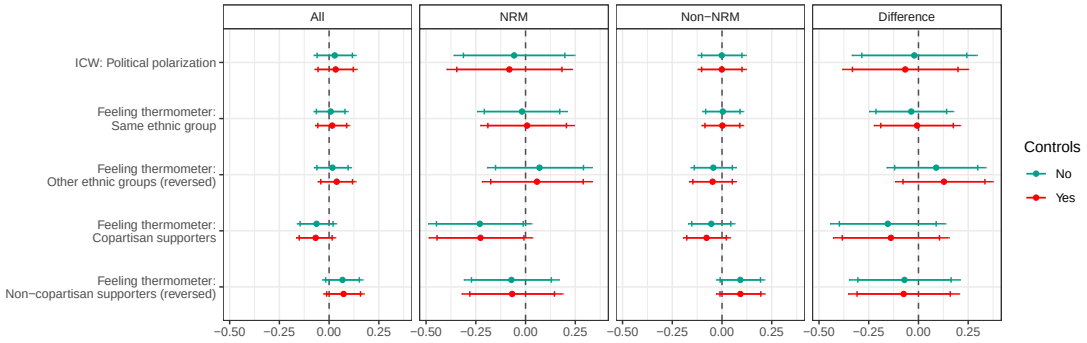
Notes: The estimates in each panel derive from equation (1) estimated in the full sample (left column) and subsamples according to NRM and non-NRM partisanship (middle and right columns). Index outcomes are standardized; subcomponents are unstandardized. The addition of controls indicates adjustment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



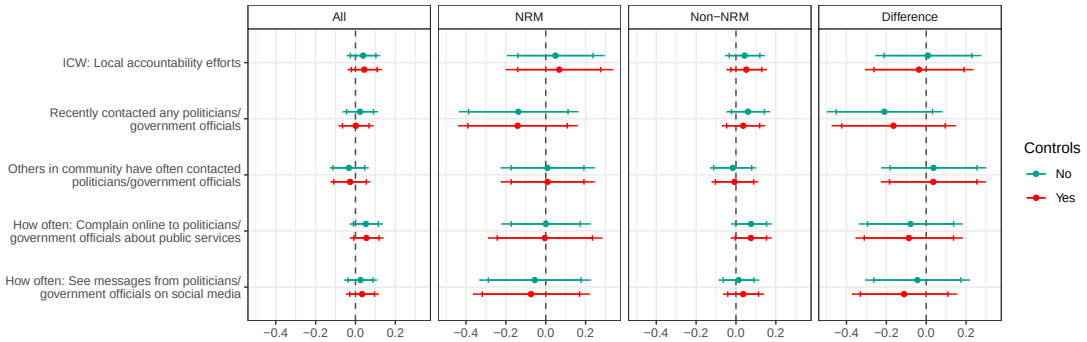
(d) Perceived state capacity



(e) Perceptions of democracy



(f) Affective political polarization

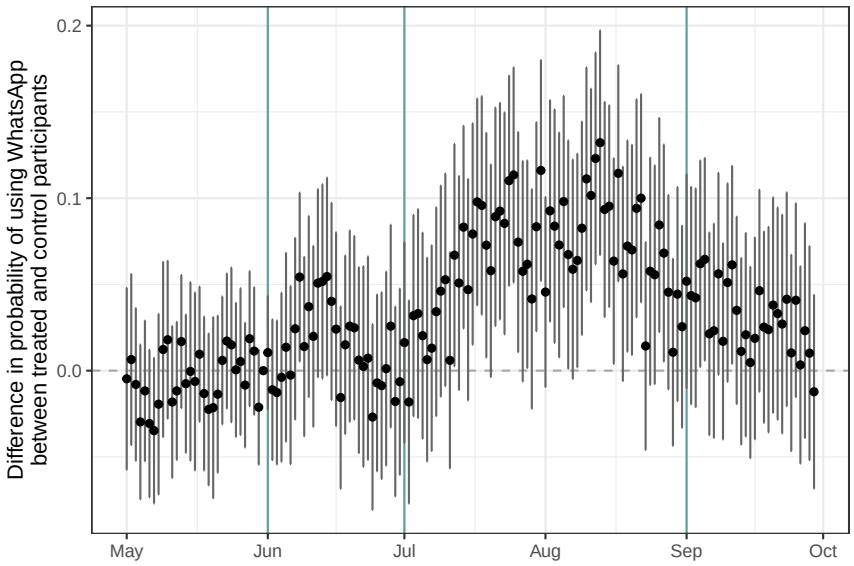


(g) Local accountability efforts

Figure A8 (cont.): Treatment effects of experimental subsidy on pre-registered indices

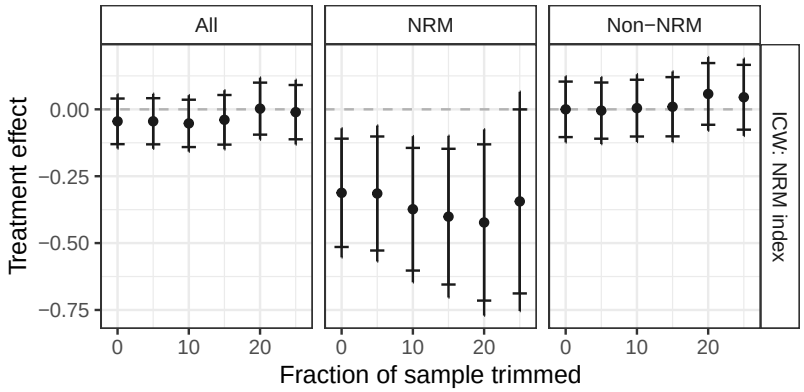
Notes: The estimates in each panel derive from equation (1) estimated in the full sample (left columns) and subsamples according to NRM and non-NRM partisanship (middle and right columns). Index outcomes are standardized; subcomponents are unstandardized. The addition of controls indicates adjustment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

Figure A8: Differences in daily use of WhatsApp, by treatment assignment



Notes: Estimates are from equation (1), where the outcome is daily use of WhatsApp per week. 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.

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.

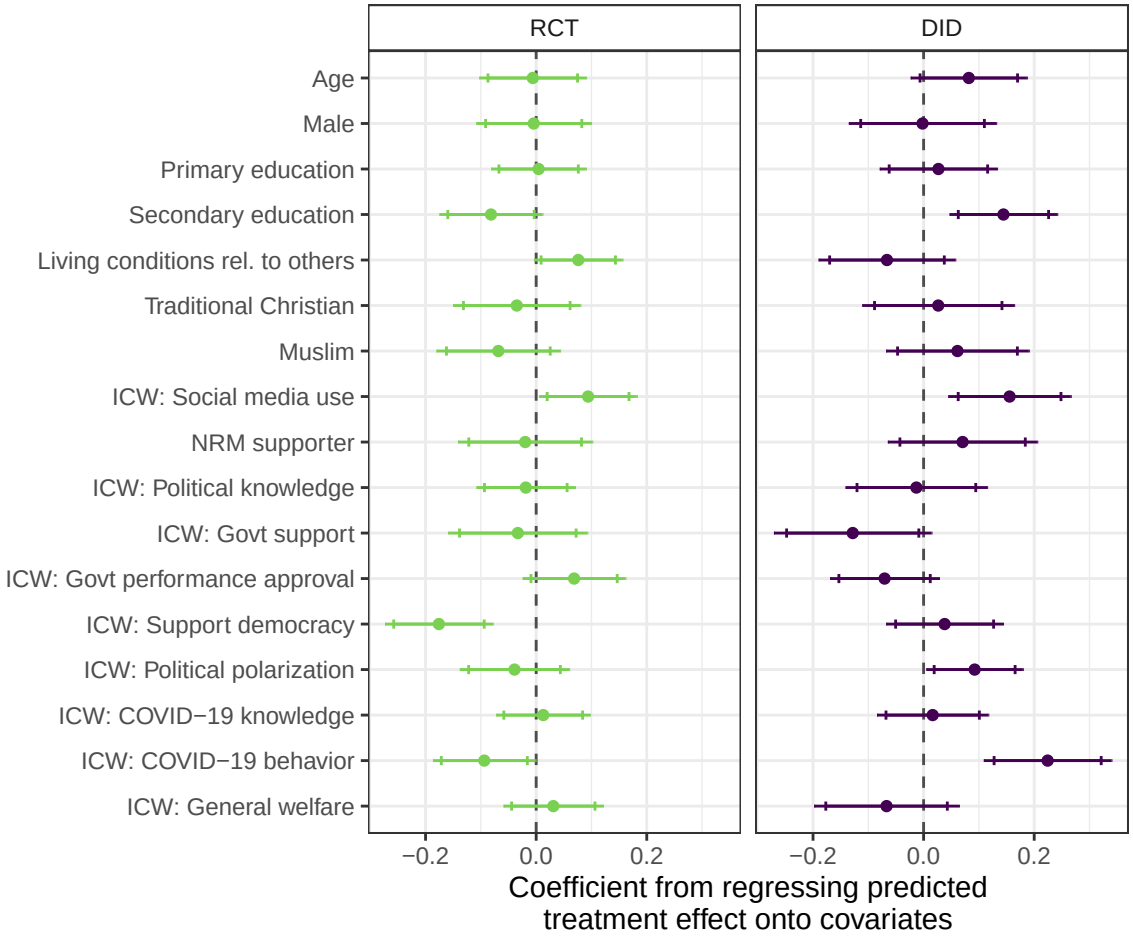
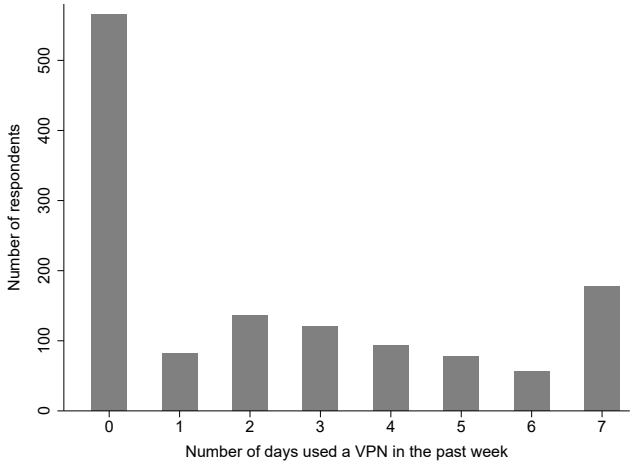


Figure A10: Correlation between baseline survey covariates and predicted WhatsApp usage in the experimental and difference-in-differences research designs

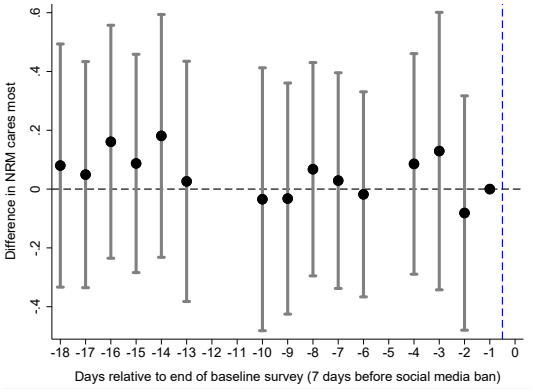
Notes: Both the RCT and difference-in-differences specifications are estimated using OLS including block fixed effects, where standardized baseline covariates differ by row. The RCT specification regresses the predicted treatment effect on WhatsApp usage from field experiment onto the set of standardized 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

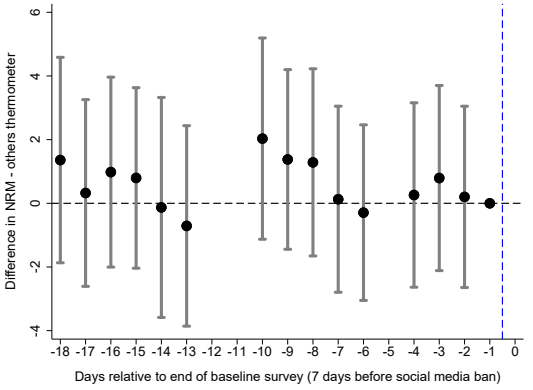


Note: Baseline survey administered in December and January 2020 asked respondents how many days they had used a VPN in the prior week.

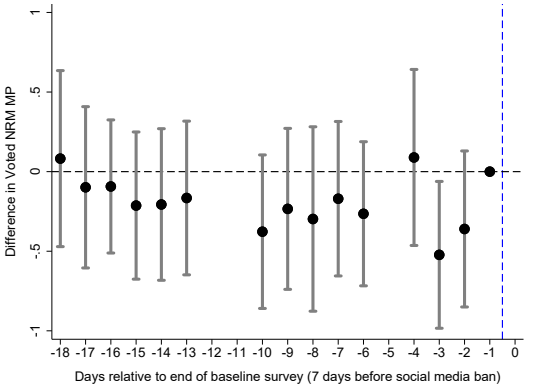
Figure A12: Event study trends in NRM support by baseline survey enumeration date



(a) NRM cares most about people like the respondent

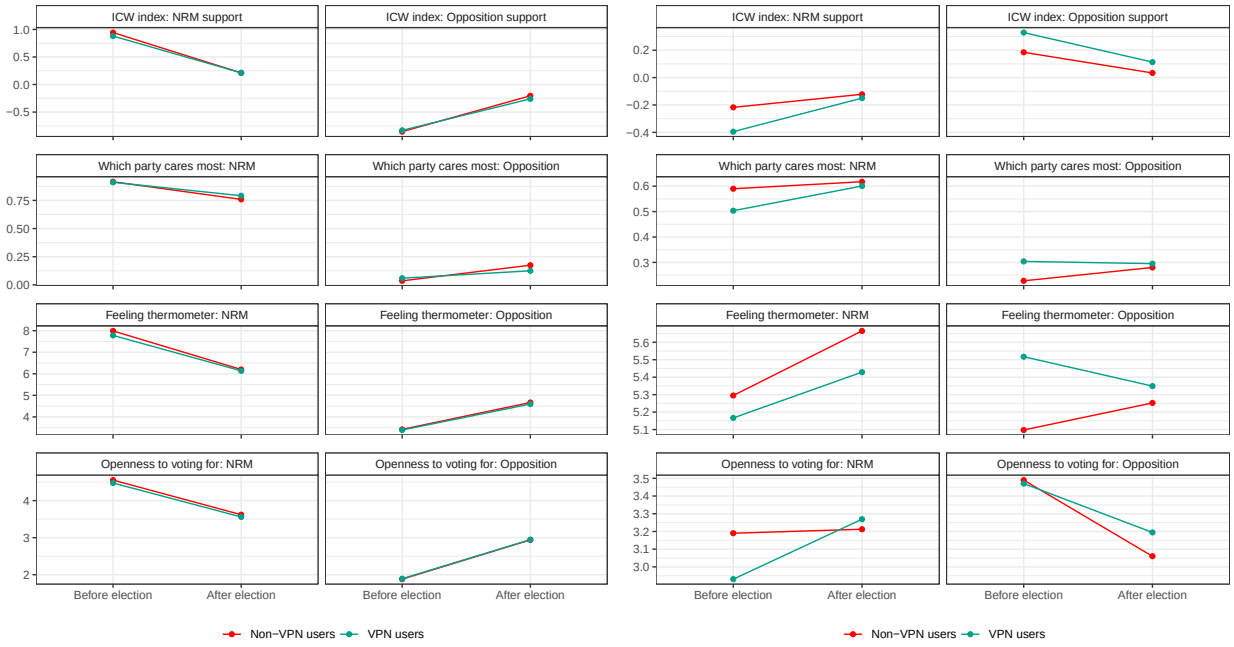


(b) NRM feeling thermometer



(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

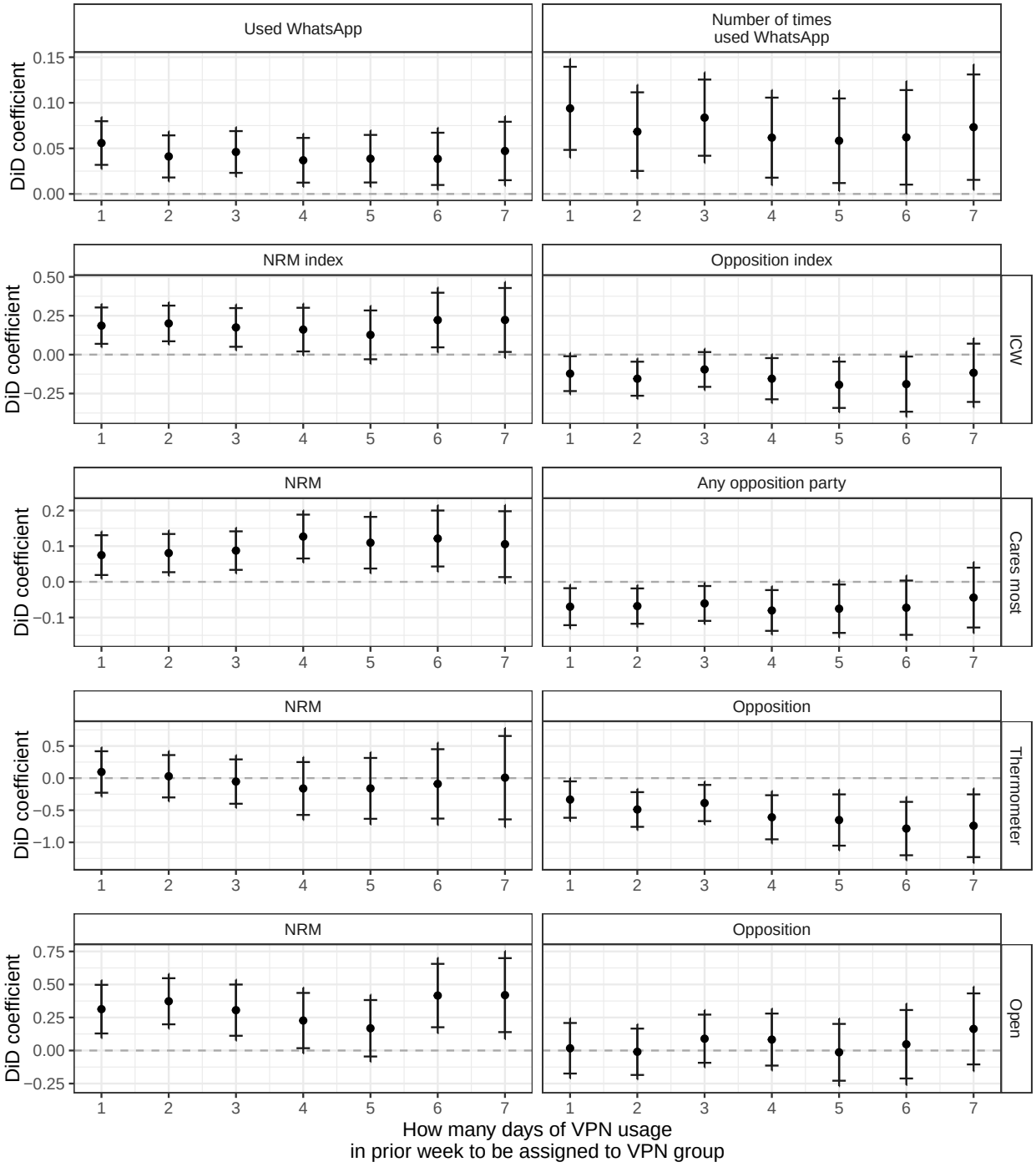


Figure A14: Varying definition of VPN group assignment based on frequency of usage

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.

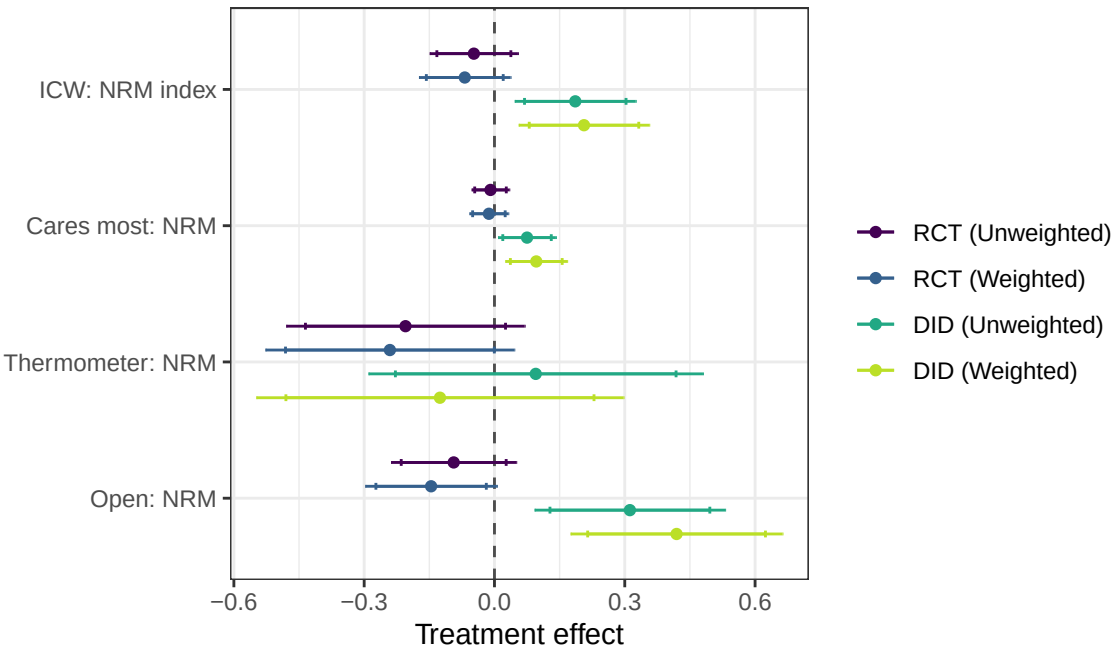


Figure A15: Reweighted estimates from experimental and difference-in-differences research designs

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 × NRM supporter			-0.020	-0.026
			(0.021)	(0.024)
Treatment + Treatment × 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 standard deviation	0.21	0.21	0.21	0.21
Block fixed effects		✓		✓

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 captures 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 captures 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	0.00	0.00	0.00	0.00
Control SD	1.00	1.00	1.00	1.00
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	0.00	0.00	0.00	0.00
Control SD	1.00	1.00	1.00	1.00

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 captures 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 captures 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 captures 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 outcome mean	0.16	0.16
Control outcome std. dev.	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**	0.198***	0.148*	0.203**	-0.118*	-0.161**	-0.058	-0.108
	(0.073)	(0.075)	(0.089)	(0.092)	(0.070)	(0.071)	(0.084)	(0.090)
VPN × Post election × NRM supporter			-0.086	-0.179			-0.020	-0.050
			(0.137)	(0.149)			(0.153)	(0.172)
VPN × Post election + VPN × Post election × NRM supporter			0.062	0.024			-0.078	-0.158
			(0.101)	(0.110)			(0.123)	(0.134)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control outcome mean	0.00	-0.00	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	1.00	1.00
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*	0.079*	-0.070**	-0.071**	-0.061	-0.067
	(0.034)	(0.035)	(0.042)	(0.043)	(0.032)	(0.033)	(0.040)	(0.042)
VPN × Post election × NRM supporter			-0.032	-0.041			-0.012	-0.004
			(0.073)	(0.083)			(0.064)	(0.074)
VPN × Post election + VPN × Post election × NRM supporter			0.038	0.038			-0.073	-0.071
			(0.058)	(0.066)			(0.046)	(0.056)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control outcome mean	0.67	0.67	0.67	0.67	0.21	0.21	0.21	0.21
Control outcome std. dev.	0.47	0.47	0.47	0.47	0.41	0.41	0.41	0.41
Trading center × Post election FEs		✓		✓		✓		✓
Panel C: Feeling thermometer (0-very cold – 10-very warm)								
VPN × Post election	0.095	0.285	-0.108	0.185	-0.334*	-0.438**	-0.323	-0.462**
	(0.196)	(0.208)	(0.235)	(0.244)	(0.172)	(0.181)	(0.209)	(0.221)
VPN × Post election × NRM supporter			0.254	-0.016			0.279	0.337
			(0.387)	(0.425)			(0.425)	(0.478)
VPN × Post election + VPN × Post election × NRM supporter			0.146	0.169			-0.045	-0.126
			(0.308)	(0.352)			(0.352)	(0.396)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control outcome mean	5.95	5.95	5.95	5.95	4.84	4.85	4.84	4.85
Control outcome std. dev.	2.67	2.67	2.67	2.67	2.50	2.50	2.50	2.50
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**	0.317**	0.017	-0.044	0.154	0.103
	(0.112)	(0.121)	(0.141)	(0.151)	(0.116)	(0.122)	(0.135)	(0.149)
VPN × Post election × NRM supporter			-0.299	-0.429**			-0.161	-0.292
			(0.205)	(0.210)			(0.223)	(0.248)
VPN × Post election + VPN × Post election × NRM supporter			0.017	-0.112			-0.007	-0.189
			(0.139)	(0.139)			(0.185)	(0.199)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control outcome mean	3.46	3.46	3.46	3.46	3.02	3.03	3.02	3.03
Control outcome std. dev.	1.40	1.40	1.40	1.40	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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Non-political support interactive covariates										
VPN × Post election	0.189**	0.195**	0.187**	0.202***	0.184**	0.200***	0.184**	0.197***	0.195***	0.205***
	(0.074)	(0.076)	(0.074)	(0.077)	(0.073)	(0.075)	(0.073)	(0.075)	(0.072)	(0.075)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control outcome mean	0.00	-0.00	0.00	-0.00	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	1.00	1.00	1.00	1.00
Trading center × Post election FEs		✓		✓		✓		✓		✓
	Age		Pol. news consumption		Pol. knowledge		Govt/NRM survey		Perceived monitoring	
Panel B: Political support interactive covariates										
VPN × Post election	0.125*	0.156**	0.109*	0.161**	0.177**	0.194**	0.134*	0.164**	0.176**	0.197***
	(0.070)	(0.072)	(0.063)	(0.066)	(0.072)	(0.075)	(0.069)	(0.070)	(0.072)	(0.075)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control outcome mean	0.00	-0.00	0.00	-0.00	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	1.00	1.00	1.00	1.00
Trading center × Post election FEs		✓		✓		✓		✓		✓
	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 pre-ban defined 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 various balancing weights

	Entropy weights				Hierarchically regularized entropy weights				Covariate balancing propensity score				Bayesian additive propensity score			
	Support for NRM		Support for Opposition		Support for NRM		Support for Opposition		Support for NRM		Support for Opposition		Support for NRM		Support for Opposition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Panel A: Support for NRM and opposition party ICW indexes																
VPN × Post election	0.196**	0.208**	-0.079	-0.113	0.192**	0.203**	-0.104	-0.148*	0.196**	0.207**	-0.079	-0.116	0.225***	0.234***	-0.104	-0.132
	(0.080)	(0.081)	(0.077)	(0.079)	(0.081)	(0.081)	(0.075)	(0.076)	(0.080)	(0.081)	(0.077)	(0.078)	(0.083)	(0.085)	(0.078)	(0.080)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control outcome mean	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	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	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Panel B: Which party cares most about people like the respondent																
VPN × Post election	0.076*	0.078**	-0.056*	-0.056	0.073*	0.074*	-0.068**	-0.067*	0.076*	0.079**	-0.056*	-0.056	0.092**	0.094**	-0.069*	-0.074**
	(0.039)	(0.039)	(0.033)	(0.035)	(0.038)	(0.038)	(0.034)	(0.035)	(0.039)	(0.039)	(0.033)	(0.034)	(0.041)	(0.042)	(0.036)	(0.036)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control outcome mean	0.67	0.67	0.21	0.21	0.67	0.67	0.21	0.21	0.67	0.67	0.21	0.21	0.67	0.67	0.21	0.21
Control outcome std. dev.	0.47	0.47	0.41	0.41	0.47	0.47	0.41	0.41	0.47	0.47	0.41	0.41	0.47	0.47	0.41	0.41
Panel C: Feeling thermometer (0-very cold – 10-very warm)																
VPN × Post election	0.156	0.277	-0.265	-0.338*	0.159	0.290	-0.309	-0.412**	0.156	0.269	-0.265	-0.348*	0.145	0.248	-0.292	-0.369*
	(0.210)	(0.222)	(0.188)	(0.196)	(0.210)	(0.222)	(0.187)	(0.193)	(0.210)	(0.222)	(0.188)	(0.195)	(0.218)	(0.235)	(0.199)	(0.208)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control outcome mean	5.95	5.95	4.84	4.85	5.95	5.95	4.84	4.85	5.95	5.95	4.84	4.85	5.95	5.95	4.84	4.85
Control outcome std. dev.	2.67	2.67	2.50	2.50	2.67	2.67	2.50	2.50	2.67	2.67	2.50	2.50	2.67	2.67	2.50	2.50
Panel D: Openness to voting for party (1-not at all – 5-very open)																
VPN × Post election	0.320***	0.289**	0.054	-0.002	0.314***	0.282**	0.039	-0.032	0.320***	0.289**	0.054	-0.004	0.367***	0.343**	0.034	0.009
	(0.117)	(0.123)	(0.124)	(0.131)	(0.119)	(0.125)	(0.120)	(0.126)	(0.117)	(0.123)	(0.124)	(0.131)	(0.130)	(0.138)	(0.127)	(0.137)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control outcome mean	3.46	3.46	3.02	3.03	3.46	3.46	3.02	3.03	3.46	3.46	3.02	3.03	3.46	3.46	3.02	3.03
Control outcome std. dev.	1.40	1.40	1.45	1.45	1.40	1.40	1.45	1.45	1.40	1.40	1.45	1.45	1.40	1.40	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: Columns 1-4: Entropy weights (Hainmueller and Xu 2013); 5-8: Hierarchically regularized entropy weights (Xu and Yang 2023); 9-12: Covariate balancing propensity score (Imai and Ratkovic 2014); 13-16: Bayesian additive regression tree propensity score (Hill, Weiss and Zhai 2011). Weights are constructed to balance survey wave 1 characteristics of VPN-using and non-VPN-using samples according to age, network, gender, education, religion, perceived living conditions, social media usage, prior political knowledge, government policy approval, support for democracy, polarization, and subjective welfare. Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A24: Differential effects of VPN use on perceptions of monitoring

	Enumerators sent by...					
	Government		NRM		Perceived monitoring index	
	(1)	(2)	(3)	(4)	(5)	(6)
VPN × Post election	0.006	0.006	-0.003	0.001	0.008	-0.027
	(0.024)	(0.026)	(0.010)	(0.010)	(0.053)	(0.055)
Observations	2,620	2,612	2,620	2,612	2,620	2,612
Control outcome mean	0.10	0.09	0.02	0.02	-0.03	-0.03
Control outcome std. dev.	0.29	0.29	0.13	0.14	0.70	0.70
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: Heterogeneity in difference-in-differences effects on NRM support, by perceptions of monitoring

	NRM support index (ICW)											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VPN × Post election	0.185**	0.199***	0.184**	0.200***	0.185**	0.197**	0.190***	0.200***	0.194***	0.205***	0.182**	0.198***
	(0.073)	(0.075)	(0.072)	(0.075)	(0.073)	(0.075)	(0.072)	(0.075)	(0.072)	(0.075)	(0.072)	(0.075)
VPN × Post election × Parties monitor vote (baseline)	0.014	0.023										
	(0.051)	(0.053)										
VPN × Post election × How careful online (baseline)			0.042	0.079								
			(0.093)	(0.100)								
VPN × Post election × Punished for online speech (midline)					-0.032	-0.062						
					(0.061)	(0.061)						
VPN × Post election × Govt. knows VPN usage (midline)							-0.053	-0.016				
							(0.057)	(0.055)				
VPN × Post election × Perceived monitoring index (all)									-0.085	-0.059		
									(0.145)	(0.144)		
VPN × Post election × Perceived monitoring index (baseline-only)											0.058	0.099
											(0.097)	(0.103)
VPN × Post election + VPN × Post election × covariate	0.199**	0.222**	0.225*	0.279**	0.153*	0.136	0.137	0.185**	0.109	0.146	0.240*	0.297**
	(0.092)	(0.092)	(0.121)	(0.129)	(0.088)	(0.090)	(0.087)	(0.086)	(0.153)	(0.155)	(0.126)	(0.129)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
Control outcome mean	0.00	-0.00	0.00	-0.00	0.00	-0.00	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	1.00	1.00	1.00	1.00	1.00	1.00
Trading center × Post election FEs		✓		✓		✓		✓		✓		✓

Notes: Each specification is estimated using OLS, and includes individual and period fixed effects. Index in columns 9-10 constructed using all four interacted variables used in columns 1-8; index in columns 11-12 constructed using only the variables observed at baseline (columns 1-4). Standard errors clustered by trading center are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A26: Changes in aggregate number of posts

	All posts		Political posts		Anti-government posts	
	(1)	(2)	(3)	(4)	(5)	(6)
Observational period × Anti-govt page	-10.23***	-9.11**	-8.71***	-7.87***	-4.99***	-4.59***
	(3.81)	(3.56)	(2.74)	(2.56)	(1.43)	(1.33)
Experimental period × Anti-govt page	-13.06***	-11.71***	-10.72***	-9.73***	-6.08***	-5.61***
	(4.38)	(4.09)	(3.15)	(2.94)	(1.64)	(1.53)
Observations	135,056	135,056	135,056	135,056	135,056	135,056
Control outcome mean	10.62	10.62	2.58	2.58	0.51	0.51
Control outcome standard deviation	47.82	47.82	18.31	18.31	7.53	7.53
Page and week fixed effects	✓	✓	✓	✓	✓	✓
Group-size decile × week fixed effects		✓		✓		✓

Notes: See Appendix A.2 for information on post classification. Each specification is estimated using OLS, and adjusts for Facebook page/group ID fixed effects and week fixed effects. Excluded time period comprises the pre-ban election campaign. Even-indexed columns interact week with decile of page/group posting frequency prior to the election campaign. Standard errors clustered by page/group in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A27: Changes in number of posts about particular topics

	Election		Regime		Economy		Services	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Observational period × Anti-govt page	-11.77***	-11.19***	-3.24**	-3.04**	0.13	0.12	0.05	0.07
	(3.70)	(3.56)	(1.30)	(1.26)	(0.11)	(0.11)	(0.11)	(0.11)
Experimental period × Anti-govt page	-14.28***	-13.71***	-4.64***	-4.45***	0.12	0.11	-0.05	-0.03
	(4.28)	(4.14)	(1.55)	(1.51)	(0.11)	(0.11)	(0.25)	(0.25)
Observations	82,059	82,059	82,059	82,059	82,059	82,059	82,059	82,059
Control outcome mean	2.68	2.68	2.78	2.78	0.56	0.56	1.54	1.54
Control outcome standard deviation	18.10	18.10	12.68	12.68	2.85	2.85	5.99	5.99
Page and week fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Group-size decile × week fixed effects		✓		✓		✓		✓

Notes: See Appendix A.2 for information on post classification. Each specification is estimated using OLS, and adjusts for Facebook page/group ID fixed effects and week fixed effects. Excluded time period comprises the pre-ban election campaign. Even-indexed columns interact week with decile of page/group posting frequency prior to the election campaign. Standard errors clustered by page/group in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A28: Changes in any posts about regime-related subjects

	General government		Security and police		Corruption		Protests and riots		Rights and freedom of speech	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Observational period × Anti-govt page	0.00	0.00	-0.05***	-0.05***	-0.01*	-0.01*	-0.05***	-0.05***	-0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Experimental period × Anti-govt page	0.00	0.00	-0.10***	-0.09***	0.00	0.00	-0.07***	-0.07***	-0.03***	-0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	82,059	82,059	82,059	82,059	82,059	82,059	82,059	82,059	82,059	82,059
Control outcome mean	0.28	0.28	0.18	0.18	0.04	0.04	0.06	0.06	0.12	0.12
Control outcome standard deviation	0.45	0.45	0.38	0.38	0.19	0.19	0.23	0.23	0.33	0.33
Page and week fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Group-size decile × week fixed effects		✓		✓		✓		✓		✓

Notes: See Appendix A.2 for information on post classification. Each specification is estimated using OLS, and adjusts for Facebook page/group ID fixed effects and week fixed effects. Excluded time period comprises the pre-ban election campaign. Even-indexed columns interact week with decile of page/group posting frequency prior to the election campaign. Standard errors clustered by page/group in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A29: Changes in any posts about election-related subjects

	Parties and candidates		Voting		Election fraud	
	(1)	(2)	(3)	(4)	(5)	(6)
Observational period × Anti-govt page	-0.03***	-0.03***	-0.07***	-0.06***	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Experimental period × Anti-govt page	-0.04***	-0.04***	-0.10***	-0.10***	-0.03***	-0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	82,059	82,059	82,059	82,059	82,059	82,059
Control outcome mean	0.29	0.29	0.18	0.18	0.02	0.02
Control outcome standard deviation	0.45	0.45	0.39	0.39	0.14	0.14
Page and week fixed effects	✓	✓	✓	✓	✓	✓
Group-size decile × week fixed effects		✓		✓		✓

Notes: See Appendix A.2 for information on post classification. Each specification is estimated using OLS, and adjusts for Facebook page/group ID fixed effects and week fixed effects. Excluded time period comprises the pre-ban election campaign. Even-indexed columns interact week with decile of page/group posting frequency prior to the election campaign. Standard errors clustered by page/group in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A30: 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	-0.172***	-0.187***								
	(0.064)	(0.070)								
Ban reduced social interactions × Post election			-0.151*	-0.163*						
			(0.089)	(0.096)						
Ban affected access to entertainment × Post election					-0.075	-0.066				
					(0.087)	(0.090)				
Ban affected purchase of goods/services × Post election							-0.086	-0.075		
							(0.130)	(0.140)		
Ban interfered with business/job × Post election									0.140*	0.079
									(0.075)	(0.076)
VPN × Post	0.181**	0.200***	0.189**	0.206***	0.187**	0.200***	0.184**	0.198***	0.182**	0.200***
	(0.072)	(0.074)	(0.073)	(0.075)	(0.072)	(0.075)	(0.073)	(0.075)	(0.072)	(0.075)
Observations	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612	2,620	2,612
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):629–676.

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.

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 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.

Hill, Jennifer, Christopher Weiss and Fuhua Zhai. 2011. "Challenges with propensity score strategies in a high-dimensional setting and a potential alternative." *Multivariate Behavioral Research* 46(3):477–513.

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.

Imai, Kosuke and Marc Ratkovic. 2014. "Covariate balancing propensity score." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 76(1):243–263.

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

Xu, Yiqing and Eddie Yang. 2023. "Hierarchically regularized entropy balancing." *Political Analysis* 31(3):457–464.

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.