MASS POLITICAL INFORMATION ON SOCIAL MEDIA: FACEBOOK ADS, ELECTORATE SATURATION, AND ELECTORAL ACCOUNTABILITY IN MEXICO*

JOSÉ RAMÓN ENRÍQUEZ† HORACIO LARREGUY‡ JOHN MARSHALL§

ALBERTO SIMPSER¶

JULY 2022

While social media has facilitated protests and challenged democratic norms, mass online campaigns could also enhance electoral accountability. Beyond increasing direct exposure to information, high saturation campaigns—that target large fractions of an electorate—may induce or amplify information diffusion, persuasion, or coordination between voters. Randomizing saturation across municipalities, we evaluate the electoral impact of non-partisan Facebook ads informing millions of Mexican citizens of municipal expenditure irregularities in 2018. The vote shares of incumbent parties that engaged in zero/negligible irregularities increased by 6-7 percentage points in directly-targeted electoral precincts and 3 percentage points in untargeted precincts within treated municipalities. These direct and spillover effects were greater where ads targeted 80%—rather than 20%—of the municipal electorate. The amplifying effects of high saturation campaigns reflect interactions between citizens within more socially-connected municipalities, rather than responses by politicians or media outlets. These findings demonstrate how mass media can ignite social interactions to promote political accountability.

*We thank Daron Acemoglu, David Broockman, Bruce Bueno de Mesquita, Eliana La Ferrara, Chappell Lawson, Ro'ee Levy, Alexey Makarin, Greg Martin, Craig McIntosh, Elias Papaioannou, Maria Petrova, Vincent Pons, Andrea Prat, Carlo Schwarz, Austin Wright, David Yang, and participants at Bocconi University, the CESS of Oxford University in Chile, EGAP, Einaudi Institute for Economics and Finance, Emory University, Harvard University, ITAM, King's College London, the Madrid Political Economy Workshop, MIT GOV/LAB, Norwegian Business School BI, the NYU-CESS Conference, Oxford University, Toulouse School of Economics, the University of Barcelona, the University of Chicago Harris School, the University of Rochester Wallis Political Economy Conference, the University of Virginia, University of Wisconsin-Madison, and Yale University for their feedback and comments in various stages of the project. We are extremely grateful to Claudia Cervantes Pérez, Ricardo de la Teja, and Christian Aarón Ramírez García for excellent research assistance and GeoPoll for their survey implementation. This research was implemented in collaboration with Borde Político, and was supported by the Dean's Competitive Fund for Promising Scholarship of the Faculty of Arts and Sciences at Harvard University and the Pershing Square Fund for Research on the Foundations of Human Behavior. The project was approved by the Columbia University Human Research Protection office (AAAR8909), the Harvard Committee on the Use of Human Subjects (18-0743), and ITAM's Institutional Review Board. The pre-analysis plan was pre-registered in the AEA RCT Registry (AEARCTR-0003135), and is publicly available at www.socialscienceregistry.org/trials/3135.

†Harvard Kennedy School, Harvard University. Email: jrenriquez@g.harvard.edu.

‡Department of Economics and Political Science, ITAM. Email: horacio.larreguy@itam.mx.

§Department of Political Science, Columbia University. Email: jm4401@columbia.edu.

¶Department of Political Science, ITAM. Email: alberto.simpser@itam.mx.

1 Introduction

Advances in digital communication technologies have created new opportunities for targeting information toward citizens *en masse* at limited cost. The most recent advances have been especially dramatic in the Global South, where the use of the internet and social media platforms—which are primarily accessed via cell phones—is rapidly catching up to levels in the Global North and grew by more than 50% between 2013 and 2018 (Poushter, Bishop and Chwe 2018). The increasing availability of these technologies is revolutionizing active and passive exposure to politically-relevant information and democratizing the supply of such information (Guriev, Melnikov and Zhuravskaya 2021; Zhuravskaya, Petrova and Enikolopov 2020).

The potential for partisan actors to manipulate or distract citizens with misinformation and government propaganda poses important challenges to democracy. However, the digital revolution also presents unprecedented opportunities to increase electoral accountability. By disseminating credible information about government performance in office, without needing to rely on under-resourced traditional media outlets that may be particularly vulnerable to political capture (Besley and Prat 2006), non-partisan actors could facilitate improved selection and control of elected representatives by voters (Fearon 1999; Ferejohn 1986). This potential for enhancing electoral accountability—the focus of this article—is particularly significant in the Global South, where politician malfeasance and low-quality public goods provision remain major challenges to development (Khemani et al. 2016).

A key feature of online communication technologies is their capacity to quickly expose many citizens to information. This can be achieved directly through investments in mass dissemination—such as ads, hosted content, or even bots—or more organically through content shared by consumers, promoted by social media platforms, or retransmitted by media outlets. In contrast with more traditional media, social media enables information distributors to engage in high saturation campaigns at low cost, where we define *saturation* as the share of an electorate targeted directly by an information campaign.[1] Mixed findings across prior studies hint that information relayed by mass media outlets may produce larger effects on voting behavior among exposed citizens than smaller-scale leaflet-, door-to-door-, or group-based information dissemination efforts (Dunning et al. 2019; Ferraz and Finan 2008). However, little is yet known about whether the level of saturation of information campaigns magnifies the effect of direct or indirect exposure to informational content. Establishing the role of saturation may have important implications for designing and regulating information campaigns.

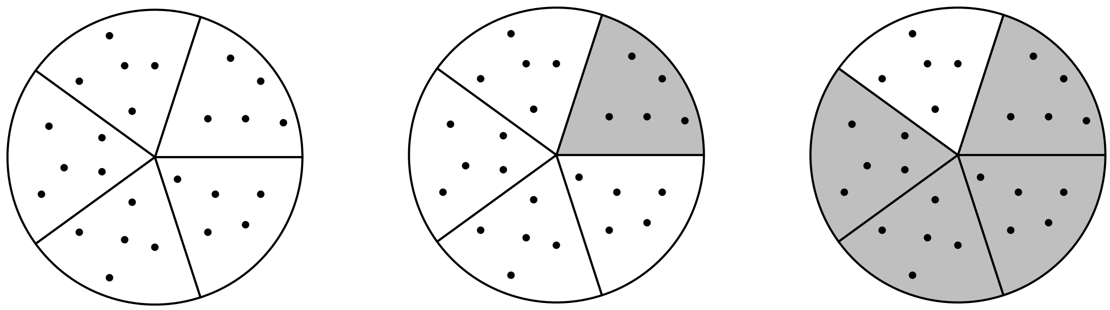
[1]Common alternative conceptions of saturation address the number of times a given piece of information is received or the point at which information could be absorbed no further. To fix terminology, this article focuses on saturation in terms of its coverage across voters within a given electorate. In this sense, we adhere to the conception of saturation as the *share* of units within a cluster that are targeted with campaign content (as in Baird et al. 2018).

Providing information at high levels of saturation could induce or amplify electoral accountability among citizens that were *and were not* targeted with information by activating social interactions between citizens. Beyond increasing exposure through diffusion by seeding information more widely within networks (Alt et al. 2022), high saturation campaigns could also trigger additional mechanisms within highly connected networks that magnify responses to direct and indirect exposure to information. Community discussion sparked by providing information about incumbent performance may shape voting behavior by enhancing belief updating about the incumbent's suitability for office, raising the salience of the discussed dimensions of incumbent performance (Iyengar and Kinder 1987), generating political persuasion by opinion leaders (Katz and Lazarsfeld 1955), or facilitating explicit or tacit voting agreements between voters (Chwe 2000; Shadmehr and Bernhardt 2017). Even without communication between citizens, high saturation campaigns could coordinate voting around public signals of incumbent performance in office by generating common knowledge about what others believe and how they are expected to vote (Cornand and Heinemann 2008; Morris and Shin 2002). Although we cannot fully pin down particular social mechanisms, our empirical strategy identifies whether saturation amplifies the impact of exposure to incumbent performance information on voting behavior.

We implemented a two-level field experiment to estimate the direct and indirect (i.e. spillover) effects of a large-scale information campaign delivered via social media, as well as variation in these effects by electorate-level *saturation*, on electoral accountability during the 2018 Mexican municipal elections. Specifically, we evaluate the impact of a non-partisan and low-cost campaign by Borde Político—a Mexican NGO promoting government transparency using digital tools. The campaign used Facebook ads to inform citizens about the extent of irregularities in municipal expenditures, which often constitute corruption (Chong et al. 2015; Larreguy, Marshall and Snyder 2020). This information was extracted from publicly-available audits conducted by the Federal Auditor's Office, Mexico's nationwide independent government auditing body, and disseminated via 26-second paid-for video ads on Facebook in the week preceding the election. Corruption was a highly salient issue during the 2018 election campaign, in which anti-establishment presidential candidate Andrés Manuel López Obrador and his recently-founded left-wing National Regeneration Movement (MORENA) party overcame traditional incumbents to win federal and local elections across the country, including the Presidency.

In collaboration with Borde Político, we first randomized whether their Facebook ad campaign documenting municipal spending irregularities targeted 0% ("control"), 20% ("low saturation"), or 80% ("high saturation") of the electorate in the 128 municipalities comprising our sample. To implement each saturation level, we divided every municipality into multiples of 5 segments (compact groups of contiguous electoral precincts). Within the treated municipalities, we then randomized the targeting of Facebook ads across segments in accordance with the assigned municipal saturation

Figure 1: Graphical illustration of randomized saturation design (treated segments within municipalities shaded in gray; black dots represent polling stations)



(a) Control municipality

(b) Low (20%) saturation municipality

(c) High (80%) saturation municipality

level: all voting-age Facebook users in 4 of every 5 segments within high saturation municipalities, and 1 of every 5 segments within low saturation municipalities, were directly targeted with Facebook ads. Since more than 70% of Mexican adults use Facebook,[2] the maximum reach of the ad campaign was closer to 14% and 56% of voters in practice because not all voters within targeted segments could be reached. This two-level randomization enables us to identify: (i) the direct effect of being targeted by the campaign within a given segment; (ii) the indirect—or "spillover"—effect of the campaign in untreated segments within treated municipalities; and (iii) how the direct and indirect effects vary with municipal saturation.[3] Given the theoretical reasons to anticipate that saturation would facilitate greater political interactions between voters, we expected the magnitude of any direct or indirect effects of the social media ad campaigns—which are likely to depend on the level of irregularities reported—to be greatest in high saturation municipalities.

Figure 1 depicts hypothetical polling stations as dots within segments and municipalities to provide intuition for our randomized saturation design. We estimate (i) by comparing treated (shaded) segments within low or high saturation municipalities with segments in control municipalities. We estimate (ii) by comparing untreated (unshaded) segments within low or high saturation municipalities with segments in control municipalities. To estimate (iii), we hold segment-level treatment assignment constant to capture saturation effects by comparing treated segments in low versus high saturation municipalities (i.e. comparing shaded segments across Figures 1b and 1c) and comparing untreated segments in low versus high saturation municipalities (i.e. comparing unshaded segments across Figures 1b and 1c).

[2] A nationally representative survey on the use and availability of information technologies (ENDUTIH) by Mexico's national statistical agency suggests that, in 2018, more than 52.3 million adults used Facebook every day and an additional 3.8 million used it every week.

[3] Studies in other domains have similarly randomized treatment across multiple levels to estimate spillover effects (e.g. Breza et al. 2021; Crépon et al. 2013; Duflo and Saez 2003; Sinclair, McConnell and Green 2012).

According to Facebook's ad campaign data, the ads ultimately reached 2.7 million unique Facebook users (appearing 3 times per person, on average) and resulted in around 15% of targeted voting-age adults—or about 20% of targeted Facebook users—watching at least 3 seconds of the ad.[4] Municipal-level engagement with the campaign was broadly proportionate with the level of access prescribed by the campaign saturation level.

Precinct-level electoral returns show that this large-scale digital information campaign substantially affected voting behavior. First, relative to the pure control segments in untreated municipalities, the best-performing incumbent parties—those whose citizens were directly informed of zero or negligible levels of irregularities—increased their vote share in the average segment that was directly targeted by Facebook ads by 6-7 percentage points, or almost half a standard deviation. The vote share of incumbent parties that presided over irregularities in the third and fourth quartiles of the irregularities distribution (i.e. those with greater levels of irregularities) was not significantly affected. Sanctioning may have been limited by the lower than normal levels of malfeasance revealed in 2018, voters already being informed about poor performance, risk-averse voters becoming less uncertain about the incumbent party's type, or swing voters already having coordinated on supporting MORENA. Furthermore, we observe spillover effects of around half this magnitude, although they are not robustly statistically significant when pooling across low and high saturation municipalities.

Second, and more importantly, we show that these direct and indirect effects are largely driven by the higher saturation information campaigns. For the 50% of mayors shown to have engaged in negligible spending irregularities, the incumbent party's vote share increased by 7-8 percentage points in treated segments within municipalities where 80% of the electorate was targeted. Furthermore, the spillover effect within such high saturation municipalities is now statistically significant and almost as large as in directly treated segments. In contrast, in municipalities where only 20% of the electorate was targeted, we observe a significantly smaller 2-3 percentage point increase in the incumbent party's vote share in treated segments and negligible indirect effects in untreated segments. Since individual levels of exposure to, and engagement with, Facebook ads were similar across treated segments in high and low saturation municipalities and we show that ad mistargeting cannot plausibly drive our estimates, these results imply a complementarity between access to Borde Político's Facebook information campaign and the share of others in the same municipality also being targeted by the campaign. In other words, high saturation bolsters both the positive direct and indirect effects of the information campaign on the vote share of the best-performing incumbent parties.

[4]Our Facebook analytics data can only distinguish whether users watched the ad at all, for at least 3 seconds, for at least 10 seconds, or entirely. Since viewers were informed of the ad's topic area at the outset and the ad allowed Facebook users to click through to access a Facebook page that showed the level of irregularities, we consider watching at least 3 seconds of an ad as the most appropriate measure of the campaign's reach.

Consistent with saturation facilitating social propagation mechanisms, we find evidence suggesting that interactions between voters caused saturation to amplify the effects of mass online information campaigns. First, the presence of substantial spillover effects suggests that factors beyond direct exposure to the ads account for most of the effect in directly- and indirectly-treated segments. Second, leveraging the probability that any two individuals within a municipality are "friends" on Facebook, we find that both the direct and indirect electoral effects of Facebook ads revealing low irregularities were greater in more socially-connected municipalities. In line with studies documenting substantial neighborhood effects on voting in Mexico (Arias et al. 2019; Finan, Seira and Simpser 2021), political discussion and coordination between voters in response to the information is likely to be greater in such municipalities. Our analyses of vote concentration further suggest that the Facebook ads coordinated support around the municipal incumbent party, although the lack of effects on vote choices in the concurrent presidential election suggests that the ads did not prime the salience of corruption as an issue. Third, non-social propagation mechanisms do not appear to be at play in our context. We find no evidence to suggest that Borde Político's ad campaign induced reactions from politicians at the end of their election campaign or media reporting on municipal malfeasance or corruption more generally.

This study makes several main contributions. First, we add to a growing literature evaluating the offline consequences of social media for political outcomes. Prior studies have found small, but cost-efficient, effects of partisan political ads on Facebook and Google on party vote shares (Hager 2019; Liberini et al. 2020). In contrast with these studies set in the Global North, we document far larger effects of *non-partisan* information—which is likely to be both more novel and more credible—on the vote shares of the best-performing incumbent parties in Mexico.[5] As efforts to regulate social media during elections grow in response to concerns about misinformation (e.g. in India and Turkey), our findings indicate that non-partisan groups can also harness social media to enhance electoral accountability. This finding buttresses studies showing that information on social media can increase turnout (Bond et al. 2012), political knowledge (Allcott et al. 2020), and protest against autocratic regimes (Enikolopov, Makarin and Petrova 2020; Steinert-Threlkeld 2017). It also provides a counterbalance to studies suggesting that social media contributes to social harms, including political polarization (Allcott et al. 2020; Fujiwara, Müller and Schwarz 2020; Levy 2021), distrust in democratic systems (Guriev, Melnikov and Zhuravskaya 2021), hate crime (Bursztyn et al. 2019; Müller and Schwarz 2021), and worse mental health (Allcott et al. 2020; Braghieri, Levy and Makarin 2021).

[5] Complementing our focus on electoral accountability, Garbiras-Diaz and Montenegro (forthcoming) similarly document large electoral effects in Colombia of non-partisan Facebook ads in an intervention conducted a year after ours. Focusing on direct effects, rather than spillover and saturation effects, they show that ads informing citizens about how to report electoral irregularities reduced electoral malpractice and then reduced the vote share of candidates likely to resort to those practices by 4 percentage points.

Second, our finding that saturation—a defining feature of broadcast, print, and especially online media—amplifies the direct and indirect effects of information dissemination helps to square the mixed evidence regarding whether credible incumbent performance information facilitates electoral accountability in developing contexts. Like this study, earlier studies examining the effects of information disseminated by media outlets (Banerjee et al. 2011; Ferraz and Finan 2008; Larreguy, Marshall and Snyder 2020; Marshall 2022) or concentrated leafleting (Arias et al. forthcoming; Chong et al. 2015) often observe significant electoral sanctions for poor performance and electoral rewards for good performance. Conversely, smaller-scale field experiments often find limited evidence that the dissemination of incumbent performance indicators affects voting behavior (Adida et al. 2020; Boas, Hidalgo and Melo 2019; Dunning et al. 2019; Humphreys and Weinstein 2012; Lierl and Holmlund 2019). Our demonstration that the direct and spillover effects of non-partisan political information increase in the local saturation of the information campaign helps to reconcile these contrasting findings.[6] While Adida et al. (2020) have also experimentally varied the saturation of an accountability campaign at the electorate level in rural Benin, the 4.5% reach of their offline intervention in the most saturation communes is substantially lower than in Borde Político's campaign.[7] At this lower level of saturation, they find that saturation amplified the effect of civics training, but not the incumbent performance information provided alongside such training. Given their mixed results, the large effects of our intervention—that reached almost 10 times more eligible voters within highly-saturation electorates—reinforce the importance of mass online information dissemination for promoting electoral accountability in the Global South.

Third, our findings contribute to the literature emphasizing that the media's impact extends beyond the direct exposure to content. Scholars of political behavior have long emphasized the "two-step flow of communication," whereby media content primarily shapes the views of citizens that regularly consume news programming but affect the mass public as these citizens pass their interpretation of the content on to others within their social network (Katz and Lazarsfeld 1955; Lazarsfeld, Berelson and Gaudet 1944). Lab experimental evidence from the US shows that the effects of partisan media spill over to individuals in discussion groups, and that these indirect effects are just as large as the direct effect when content is relayed within ideologically-congenial groups (Druckman, Levendusky and McLain 2018). Outside controlled environments, Yanagizawa-Drott (2014) shows that indirect effects of exposure to "hate radio" in Rwanda on militia violence were at least as large as the direct effects. Caprettini et al. (2021) also find that Nazi propaganda had both

[6] Appendix A.1 describes how prior studies vary in terms of information campaign saturation. None of these studies exogenously varied high degrees of saturation. The main exception to the correlation between information campaign saturation and effect magnitude is Bhandari, Larreguy and Marshall (forthcoming), where information diffusion that resulted from a small-scale campaign was substantial in rural Senegal.

[7] Buntaine et al. (2018) have also experimentally varied the *village* level saturation of a similar accountability campaign in Uganda, but do not vary saturation at the *electorate* level as we do and—perhaps unsurprisingly—fail to detect differential saturation effects.

direct and indirect effects, and that social spillovers are of similar importance as direct exposure. We add to this literature by showing that a similar logic applies to non-partisan information, using a field experiment specifically designed to distinguish direct and indirect effects of information and differential effects by information saturation.

The article is structured as follows. Sections 2 and 3 describe the context and experimental design. Section 4 then describes the campaign's reach, before we report our main results in Section 5 and explore mechanisms in Section 6. Section 7 concludes.

2 Mayoral malfeasance and accountability in Mexico

Mexico's 2,463 municipal governments are led by mayors typically elected to three-year terms, which became renewable in most states in 2018. These governments are responsible for delivering basic public services and managing local infrastructure, which can—if used effectively—play an important role in poverty alleviation and local development (Rodríguez-Castelán, Cadena and Moreno 2018). However, municipal government malfeasance remains common.

2.1 Independent audits of municipal social infrastructure spending

An important source of funding for mayors is the Municipal Fund for Social Infrastructure (FISM). These direct federal transfers represent around a quarter of the average municipality's budget and must be spent on infrastructure projects that benefit (i) localities deemed to be marginalized by the National Population Council (CONAPO), (ii) citizens in extreme poverty, or (iii) priority zones.[8] In 2010, the CONAPO defined 79% of localities as marginalized. Eligible projects include investments in the water supply, drainage, electrification, health infrastructure, education infrastructure, housing, and roads.

In the period under study, Mexico's independent Federal Auditor's Office (ASF) audited the use of FISM funds in 150-200 municipalities each year. The annual number of audits has since been reduced. ASF audits are announced after spending has occurred, and address the spending, accounting, and management of FISM transfers from the previous fiscal year. The ASF selects municipalities on the basis of the importance of FISM transfers to the municipal budget, historical performance, factors that raise the likelihood of irregularities in the management of funds, and whether the municipality has recently been audited (see Auditoría Superior de la Federación 2014). The large municipalities comprising most of the country's population have now received multiple audits since systematic audits began in 2004.

This article focuses on irregularities in FISM expenditures. Irregularities typically entail funds

[8]Localities are the smallest geographical units recognized by Mexico's national statistical agency.

that were spent on projects not benefiting marginalized localities (based on the distribution criteria above), spent on unauthorized projects that did not constitute social infrastructure projects (e.g. personal expenses and election campaigns), or were unaccounted for in the municipal budget. The audit reports indicate that such irregularities often arise from failing to demonstrate that a project benefited its intended recipients, transferring funds to non-FISM bank accounts or contractors, or failing to produce documentation proving that expenses related to claimed projects. These actions can constitute corruption in the form of kickbacks, preferential contracting, and embezzlement. Between 2009 and 2018, the ASF determined that 17% of FISM expenditures were subject to irregularities.[9]

The potential for voters to punish high levels of mayoral malfeasance and reward clean incumbents is limited by an electorate largely uninformed about the ASF's reports. Most citizens are unaware of the resources available to mayors and even their responsibility to provide basic public services in the first place (Chong et al. 2015). The ASF's reports are publicized in some media outlets and have been shown to influence voting behavior in urban environments (Larreguy, Marshall and Snyder 2020). However, because media coverage is not widespread and voter engagement with news programming varies, further dissemination of such information has the potential to significantly alter voters' beliefs and voting behavior. Indeed, Arias et al. (forthcoming) find that distributing the results of ASF reports via non-partisan leaflets caused voters to update their high expectations of incumbent party malfeasance, and in turn vote for incumbent parties. Chong et al. (2015) have also found that publicizing severe levels of unauthorized FISM spending could breed voter disengagement, with a particularly detrimental effect on support for challenger parties. This article advances beyond these prior studies by investigating whether the mass provision of information via social media, and the extent of its saturation, facilitate electoral accountability.

2.2 Electoral context

Until recently, electoral competition in Mexican municipalities was generally between two of the country's main three parties. In most parts of the country, the formerly dominant (yet now much diminished) PRI competed against either the right-wing National Action Party (PAN) or the PRI's left-wing offshoot Party of the Democratic Revolution (PRD). In 2014, ex-PRD leader Andrés Manuel López Obrador formed MORENA, a new left-wing and anti-corruption party which stood for the first time in 2015. Although MORENA's local presence was limited in 2015, it swept

[9]Given that other programs and non-federal transfers are not subject to such audits, mayoral malfeasance could be greater on other dimensions. However, we expect malfeasance across areas to be positively correlated, and thus that information about irregularities in FISM expenditures may be indicative of an administration's broader malfeasance. For example, the correlation between the irregularities in the expenditure of FISM and FORTAMUN (Fund for Municipal Strengthening) resources for the sample of municipalities whose 2016 expenditure for both federal funds was audited is 0.62.

the 2018 elections as López Obrador's message of change won a landslide presidential election. MORENA's national success extended to local elections as well, with MORENA claiming multiple governorships and hundreds of mayoral offices across the country. The 2018 election campaign was unusual in the emphasis on reducing corruption and the limited partisan attachments that enabled the overwhelming success of a party that had previously held few legislative or executive offices.

For several reasons, municipal election campaigns in Mexico are generally oriented around political parties, rather than specific candidates. First, given that consecutive re-election for mayors was only permitted for the first time in 2018, citizens are better informed about parties than individual politicians (e.g. Arias et al. forthcoming; Chong et al. 2015; Larreguy, Marshall and Snyder 2020). Second, voters may recognize that Mexico's main parties use distinct candidate selection mechanisms that select candidates with similar characteristics over time (Langston 2003). For these reasons, voters have generally held parties responsible for the actions of individual politicians (e.g. Chong et al. 2015; De La O 2013; Larreguy, Marshall and Snyder 2020; Marshall 2022). Although only 22% of mayors sought re-election in 2018, there is thus good reason to believe that voters would draw inferences about the party of the mayor whose audited expenditures were publicized by Borde Político's campaign when deciding how to vote.

2.3 Political information and the social media environment

While broadcast media outlets have traditionally been the primary source of political information in Mexico, mobile technology and social media have created new opportunities for information dissemination. According to Hootsuite and We are Social (2018*a*,*b*), 65% of Mexicans accessed the internet in 2018, with the average respondent spending more than eight hours a day online—the 7th highest rate in the world. Moreover, in 2018, 72% of adults owned a smartphone—the primary means through which adults access the internet in Mexico—and 64% of adults used social media; social media users reported spending an average of more than three hours a day using it. With almost all social media users using Facebook at least once a month, Mexico ranked 5th in the world in terms of active Facebook users.

Growing access to digital information has emerged alongside substantial amounts of credible and fake political information disseminated through social and traditional media. Due to their popularity among Mexican citizens, Facebook and WhatsApp are the prime channels for spreading political content in the form of videos, images, and memes. Misinformation was a particular concern during the 2018 election campaign, where political parties were accused of disseminating misinformation online to influence voter behavior.[10] Several Facebook pages that were identified

[10]It is likely that bots were used to exploit Facebook's algorithm to increase the visibility of certain posts, pages, and ads. For example, Facebook pages criticizing López Obrador featured posts with thousands of "likes," but no other reactions or comments, suggesting the work of bots; see www.reuters.com/article/us-mexico-facebook/in-mexico-fake-

as the most prolific misinformation distributors had between one and two million followers at the time of the election. The attacks were largely directed against López Obrador.[11] However, other candidates across all races were also affected by similar types of attacks.

3 Experimental design

We designed a field experiment to evaluate whether non-partisan dissemination of incumbent performance indicators via Facebook ads promotes electoral accountability, and the extent to which any direct or indirect effects are amplified by an information campaign's saturation. We partnered with Borde Político—a Mexican NGO that uses digital technologies to promote government transparency across the country—to evaluate the impact of their online accountability campaign ahead of Mexico's July 1, 2018 elections. Borde Político's Facebook ad campaign, which focused on the municipal elections, targeted voting-age adults and randomly varied the share of segments within different municipalities that were targeted with information about the FISM program and about the share of audited resources that were subject to irregularities in the municipality. This section describes the treatment conditions, sample, two-level randomization, measurement of outcomes, and pre-specified estimation strategy; we conclude by discussing ethical considerations.

3.1 Treatment conditions

Borde Político's Facebook ad campaign reported the results of the ASF's audit in a given municipality.[12] Figure 2 shows the slides constituting each 26-second video; music played in the background for users that unmuted the video. The first screen was designed—based on initial pilots—to attract viewers. The video next informed viewers that the FISM program transfers federal funds to municipalities for social infrastructure projects benefiting the poor. The following screens then informed viewers of how much money their municipal government received, and the percentage of the audited funds that were subject to irregularities.[13] The ads were accompanied by a legend indicating that Borde Político is a non-partisan NGO that aims to inform citizens and included links to the Borde Político and ASF websites. Users could click to access the municipality-specific Facebook page that promoted each ad. These pages included a cover photo highlighting the FISM funds re-

news-creators-up-their-game-ahead-of-election-idUSKBN1JO2VG for more details.

[11]For example, fake pictures of rallies with very few attendees were circulated to claim that López Obrador's support was deflating (see www.verificado.mx/imagenes-erroneas-eventos-amlo for more details).

[12]While the nature of the content provided is similar to prior interventions in Mexico (Arias et al. forthcoming; Chong et al. 2015), this study differs by focusing on the impact of digital dissemination and municipal campaign saturation.

[13]Incumbent performance was not benchmarked against other municipalities or previous municipal governments because prior studies detected no effect of providing additional information from comparable Mexican municipalities (Arias et al. 2018) and because some municipalities had not previously experienced an ASF audit.

ceived and the fraction of expenditures that were subject to irregularities, as well as an infographic reporting this information in greater detail (see Appendix Figures A1a or A1b for examples).

Each municipality ad campaign ran for a week, concluding on June 27, 2018—the last day of official campaigning. Incumbents, therefore, could not alter their performance in office in response to the ads and had very limited time to respond during the election campaign. Moreover, a majority of respondents in a parallel survey regarded the ad's content as somewhat or very credible; see Section A.5.2 for a discussion of the survey and its limitations.

The saturation of Borde Político's Facebook ad campaign varied across municipalities. In low saturation municipalities, Facebook ads were geographically targeted at 20% of the voting-age (18+) municipal population. In high saturation municipalities, 80% of this population was targeted.[14] The 20% and 80% saturation levels were chosen to capture a large difference in saturation that could plausibly alter the extent of social interactions within municipalities, while also maximizing statistical power.[15] The average municipal ad campaign cost around US$200, representing around 10% of a typical municipal election campaign's budget.[16]

The Facebook ad campaigns could not be designed to ensure that ads would reach all voting-age adult Facebook users within targeted locations a specific number of times. Rather, for a given payment, Facebook allows the purchaser of ads to specify the budget for each campaign and indicates the expected reach (in terms of number of users) within a geographic area. Consequently, in both the low and high saturation municipalities, Borde Político's ad campaigns were funded to the point of being able to reach the maximum possible voting-age population in each directly treated segment. While Facebook does not publicly disclose its technology for identifying user locations, Facebook's Data For Good datasets note that locations in 2020 were defined by where individuals spend most of their time at night.[17] This location is likely to coincide with the residence where voters are registered to vote, but we conduct a number of robustness checks to show that our main findings are unlikely to be driven by ad mistargeting.

The ad's content was also randomized to subtly vary common knowledge about the ad campaign's reach. Facebook users in some locations were informed at the end of the ad that the ad campaign could reach 20% or 80% of citizens in their municipality.[18] We ultimately observe no

[14] In all treated (and some control) municipalities, WhatsApp messages were sent to a mean of 50 surveyed registered voters as part of a concurrent panel survey seeking to understand the mechanisms underlying the Facebook campaign. Since only 0.02% of registered voters received the messages, we disregard them when defining municipal treatments. We find no significant difference in electoral outcomes across control municipalities that did and did not contain respondents who received WhatsApp messages.

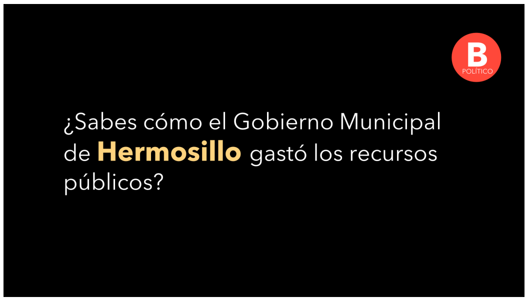
[15] Following Baird et al. (2018), we selected saturation levels to approximately minimize the equally-weighted sum of the standard errors for direct and indirect effects.

[16] These figures are from the 2018 spending reports of municipal election campaigns, which are self-reported by candidates and audited by the National Electoral Institute.

[17] When a user's GPS data is unavailable, our conversations with Facebook staff indicate that locations are based on data including the user's IP address, search traffic, and the locations of a user's friends.

[18] This (accurate) information was conveyed by adding the screen in Appendix Figure A5 to the end of the video.

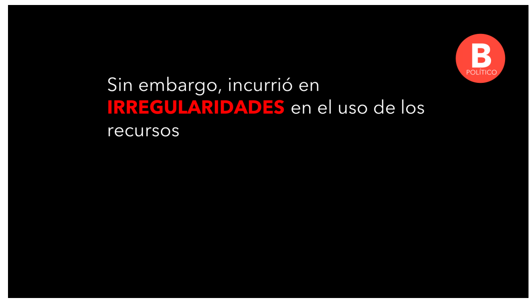
Figure 2: Example of the slides included in the ad video (from Hermosillo, Sonora)



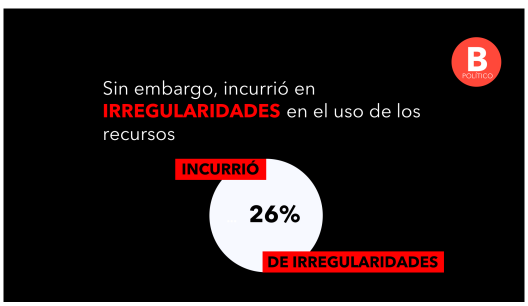
(a) Slide 1 (4 seconds)

(b) Slide 2 (4 seconds)

(c) Slide 3 (5 seconds)

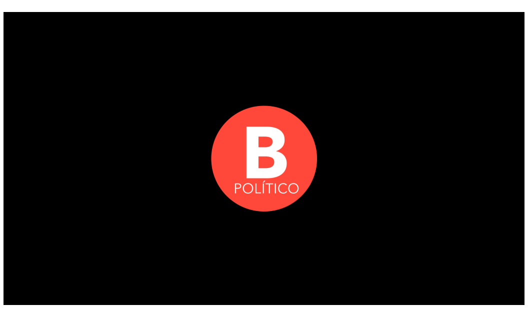


(d) Slide 4 (4 seconds)



(e) Slide 5 (4 seconds)

(f) Slide 6 (4 seconds)



(g) Slide 7 (1 second)

Note: In English: slide 1 says "Do you know how the municipal government of Hermosillo spent public monies?"; slide 2 says "In 2016, the municipal government of Hermosillo received funds from the Fund for Social Infrastructure"; slide 3 adds "It received $65 million for infrastructure"; slide 4 says "However, it incurred in irregularities in the spending of the funds"; slide 5 adds "It incurred in 26% of irregularities"; and slide 6 says "Unauthorized spending and targeting people other than the intended beneficiaries are irregularities that cause damage to government finances".

discernible differences in viewership of, or reactions to, the ads with and without explicit common knowledge communication (see Appendix Table A12). Since this treatment variant also did not differentially affect voting behavior (see Appendix Table A13), we henceforth pool the Facebook ads with and without common knowledge in all analyses.

3.2 Sample of municipalities

In 2017 and 2018, the ASF released audit reports pertaining to FISM expenditures in 561 municipalities. Of these, 128 municipalities satisfied the two criteria for inclusion in this study: (i) being located in one of the 17 states that held municipal elections in 2018; and (ii) the mayor in office before the election also being the mayor that presided over the audited expenditures.[19] These municipalities are shown in Appendix Figure A2. Collectively containing around 30 million people, roughly a quarter of Mexico's population, this sample is broadly nationally representative.

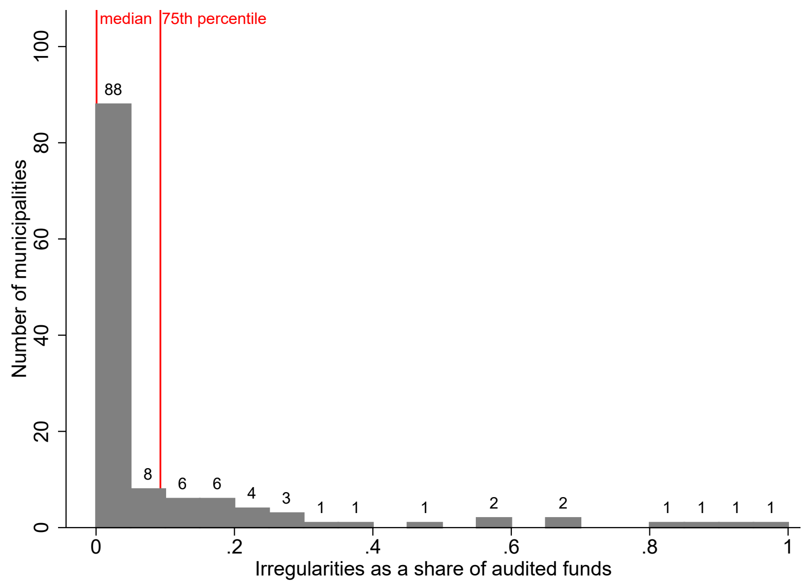
The majority of ASF audits in these municipalities reported irregularities well below 10%. As Figure 3 illustrates, no irregular spending was found in 61 of our 128 municipalities, while irregularities were just 0.07% in the median municipality. The mean share of irregular spending across municipalities was 9.2%, due to the positive skew driven by several egregious cases. Given the public's low expectations of politicians in Mexico (Arias et al. forthcoming) and lower levels of FISM malfeasance detected in 2018 in comparison with previous years, the ads likely reported better performance than many citizens expected.

3.3 Municipal- and segment-level randomizations

To evaluate the electoral impacts of Borde Político's Facebook ad campaign, we designed a two-level randomization strategy. We first assigned campaign saturation at the municipality level and then assigned Facebook ads to geographic segments (defined below) within the municipalities that received a non-zero campaign saturation. This design enables us to identify the effect of information provision via Facebook ads on voting behavior in treated segments (direct effects), the extent to which such information provision affected non-treated segments within partially treated municipalities (indirect effects), and whether these two effects vary by municipal saturation level (differential effects by saturation). Figure 1 summarizes the design graphically.

[19]These states are: Baja California Sur, Campeche, Chiapas, Colima, Estado de México, Guanajuato, Guerrero, Jalisco, Michoacán, Morelos, Nuevo León, Puebla, Querétaro, San Luis Potosí, Sonora, Tabasco, and Yucatán. Municipalities from states like Coahuila, where mayors were elected in 2017 and thus were not responsible for the spending audited by the ASF, were not included in our sample. An additional 7 states held municipal elections, but none of these municipalities were eligible for Borde Político's campaign due to their shorter electoral cycles. Other 7 states did not hold municipal elections. We excluded audited delegaciones in Mexico City because these administrative units operate differently from municipalities.

Figure 3: Distribution of irregularities across municipalities in our sample



Notes: The width of each bin is 0.05. The 25th, 50th, and 75th percentiles of the distribution are, respectively, 0%, 0.07%, and 9.3%. The mean is 9.4%, and 61 municipalities registered exactly 0% irregularities.

For the municipal level randomization, each municipality was assigned to one of 42 blocks containing 3 municipalities governed by the same incumbent party on the basis of the Mahalanobis distance over 28 covariates. For simplicity, we exclude the rump block containing 2 municipalities from our analysis.[20] Within each block, one municipality was assigned to each of the following conditions:

1. *Control*: no Facebook ads;
2. *Low saturation information campaign*: Facebook ads were geographically targeted at all adults in 1 out of every 5 segments within the municipality; and
3. *High saturation information campaign*: Facebook ads were geographically targeted at all adults in 4 out of every 5 segments within the municipality.

Since around 70% of Mexican adults regularly use Facebook, the average targeted share of registered voters was effectively 14% in low saturation municipalities and 56% in high saturation municipalities. Our blocking procedure ensured that each municipality had an equal probability of being treated, without differentially targeting treatment toward incumbents from any particular political party. Appendix Table A4 shows that the campaign saturation conditions are well balanced

[20]This deviation from our pre-analysis plan simplifies estimation by maintaining a constant probability of treatment assignment across municipalities. Appendix Table A3 reports similar results when the two rump municipalities are included.

across 52 predetermined municipal-level covariates, including incumbent party vote share at the previous election.

Within treated municipalities, we then randomized the targeting of Facebook ads to clusters of electoral precincts—Mexico's smallest geographical electoral unit. To target up to 20% and 80% of adults in low and high saturation cases, we divided each municipality into multiples of 5 evenly-populated "segments." In most municipalities, we created five segments. In larger municipalities, where it was feasible to target more segments using Facebook's targeting system, we created multiples of five segments. The resulting 783 segments were constructed from contiguous electoral precincts to form compact polygons with similar populations of individuals aged 18 or above (according to the 2010 Census).[21] Complete randomization was used to assign one in five segments within low saturation municipalities, and four in five segments within high saturation municipalities, to be targeted with Borde Político's Facebook ads. As Appendix Table A5 shows, the segment-level treatment conditions are well balanced across 63 predetermined segment-level covariates.

3.4 Measurement of primary outcomes

Our main outcomes are taken from the precinct-level electoral returns collated by Mexico's state electoral institutes. Given Mexico's party-centric political system, we focus primarily on the municipal incumbent party's vote share, as a share of total votes cast.[22] Electoral returns were available in all but two municipalities—Oxchuc, Chiapas and Ayutla de los Libres, Guerrero—which, in 2018, adopted a customary system for selecting their mayors that did not involve direct election. After excluding the block containing the 2 rump municipalities, this yielded a final sample of 124 municipalities containing 773 segments. A PRI mayor was the incumbent in 45 of these municipalities, while 39 municipalities had PAN mayors, 15 had PRD mayors, 10 had PVEM mayors, 2 had MORENA mayors, and the remainder were governed by small parties or an independent. We introduce additional data used to measure engagement with the treatments and potential mechanisms as it becomes relevant for our analysis.

[21] These segments were first generated to maximize contiguity, compactness, and population equality using the freely-downloadable program Auto-Redistrict (http://autoredistrict.org). We then manually adjusted precinct allocations to smooth the edges of segment polygons in order to facilitate the targeting of Facebook ads, given that Facebook's target markets are defined by the union of a sequence of points each with a 1km radius. The total number of segments is not a multiple of 5 because one small municipality contained only 3 electoral precincts.

[22] We obtain similar results when the vote share denominator is the predetermined number of registered voters (see Appendix Table A15).

3.5 Estimation

Our pre-registered specifications leverage the multiple layers of randomization to identify various effects of Borde Político's Facebook ad campaign on precinct-level electoral results.[23] First, we follow Baird et al. (2018) in leveraging the segment-level ad targeting to estimate direct and indirect average treatment effects using the following specification:

$$Y_{psm} = \alpha Y_{psm}^{lag} + \beta Facebook\ ads_{sm} + \gamma Spillover_{sm} + \mu_b + \varepsilon_{psm}, \tag{1}$$

where Y_{psm} is an outcome in precinct p within segment s of municipality m, Y_{psm}^{lag} is a predetermined 2015 election outcome, $Facebook\ ads_{sm}$ is an indicator for segments directly targeted with Facebook ads, $Spillover_{sm}$ is an indicator for segments that were not directly targeted with ads but are located within treated municipalities, and μ_b are fixed effects for the municipal-level randomization block. Following McKenzie (2012), we include the lagged outcome to increase estimate precision; Appendix Table A9 reports similar results without this pre-specified covariate. Standard errors are clustered by municipality to reflect the higher level of randomization at the municipal level.

Throughout, observations are weighted by: (i) the inverse probability of treatment assignment; and (ii) each precinct's share of the segment's 2010 adult population aged 18 or above. The first weight follows Horvitz and Thompson (1952) in equalizing effective sample sizes across treatment conditions, while the second ensures that each segment is weighted equally. The coefficients β and γ then respectively capture the intent to treat effects of being directly targeted by the Facebook ad campaign and being located in a non-targeted segment within a treated municipality, both relative to our reference category—segments in control municipalities that received no Facebook ads.

Second, we estimate the differential effects of the Facebook ad treatments across municipal saturation levels using OLS regressions of the following form:

$$\begin{aligned} Y_{psm} = \; & \alpha Y_{psm}^{lag} + \beta_1 Facebook\ ads\ in\ Low\ Saturation_{sm} + \beta_2 Facebook\ ads\ in\ High\ Saturation_{sm} \\ & + \gamma_1 Spillover\ in\ Low\ Saturation_{sm} + \gamma_2 Spillover\ in\ High\ Saturation_{sm} + \mu_b + \varepsilon_{psm}, \end{aligned} \tag{2}$$

where $\beta_2 - \beta_1$ and $\gamma_2 - \gamma_1$ respectively capture the differential direct and indirect effects of the Facebook ad treatments attributable to the high rather than the low saturation campaign. If saturation amplifies any effects of Facebook ads, we expect $\beta_2 - \beta_1 > (<) 0$ and $\gamma_2 - \gamma_1 > (<) 0$ when the average treatment effects are positive (negative).

Since the level of irregularities differs across municipalities, we further examine how treatment effects vary with the content of the ads. It is not obvious *a priori* how information provision will

[23] One deviation from the pre-analysis plan (excluding the rump block) and one clarification to the pre-analysis plan (operationalizing reported irregularities for heterogeneous effect analyses) are explained in Appendix A.5.1.

affect support for the incumbent on average, because this is likely to depend on how information relates to voters' prior beliefs (Ferraz and Finan 2008). While examining heterogeneity by information content was pre-specified, how the level of reported irregularities would be operationalized was not. We follow Cavalcanti, Daniele and Galletta (2018) and Larreguy, Marshall and Snyder (2020) in adopting two non-parametric approaches that split municipalities into bins where reported irregularities are likely to exceed and fall below citizens' expectations. First, we divide municipalities between those above and below the median level of irregularities (0.07%). Second, we divide the municipal distribution of reported irregularities into quartiles. The bottom two quartiles (henceforth Q1 and Q2) are pooled because, as Figure 3 shows, 48% of municipalities registered zero irregularities. Irregularities in Q3 range from 0.08% to 7.4% of audited FISM funds, with a mean of 2.2%. Irregularities in Q4 range from 8.5% to 100%, with a mean of 31.7%.

We estimate the conditional average treatment effect in each bin by extending equations (1) and (2) to include interactions between treatment conditions and a municipality's position in the irregularities distribution. For example, the interactive version of equation (1) entails estimating:

$$\begin{aligned} Y_{psm} = {} & \alpha Y_{psm}^{lag} + \beta_1 Facebook\ ads_{sm} + \beta_2 \big(Facebook\ ads_{sm} \times \mathbf{X}_m\big) \\ & + \gamma_1 Spillover_{sm} + \gamma_2 \big(Spillover_{sm} \times \mathbf{X}_m\big) + \delta \mathbf{X}_m + \mu_b + \varepsilon_{psm}, \end{aligned} \tag{3}$$

where $\mathbf{X}_m$ is either an indicator for municipalities where irregularities exceeded the sample median or a vector containing indicators for municipalities in Q3 or Q4 of the sample irregularities distribution. In each case, β_1 and γ_1 capture the direct and indirect effects of treatment in the below-median reference group; analogous conditional average treatment effects in the above-median, Q3, and Q4 subsamples are obtained by adding the corresponding interaction coefficient. Given the distribution of malfeasance, we expect to observe increases in incumbent party support in Q1/2 and decreases incumbent party support in Q4.

Two issues arise when interpreting how the effects of Facebook ads vary with the content of such ads. First, the identification of conditional average treatment effects requires the randomization to also hold within each subgroup. The balance tests in Appendix Tables A6 and A7, which report few significant differences in predetermined covariates across treated, spillover, and control segments above and below the median and within each quartile, support this assumption. Second, whereas information provision and saturation were randomly assigned, differences in treatment effects by the level of irregularities reported may not necessarily be attributable to receiving this content because ASF audit outcomes could not be randomly assigned. We address this issue by adjusting for potential differences across municipalities using interactions between treatment and a variety of potential confounders that we describe in Section 5.5.

3.6 Ethical considerations

Our collaboration with the non-partisan NGO Borde Político followed prevailing ethical and legal standards. First, the study was approved by the Institutional Review Boards at the universities of all authors. Second, the intervention also complied with Mexican electoral law. Mexican electoral authorities and a local electoral lawyer indicated that electoral law permits NGOs to exercise the freedom of expression they enjoy as collectives of citizens in order to disseminate non-partisan information about municipal government performance.

Beyond satisfying institutional and legal requirements, we regard our evaluation of Borde Político's information campaign as both ethical and academically valuable for a number of reasons. First, the intervention evaluates the impact of information provided on behalf of a non-partisan NGO that frequently disseminates politician performance information online, including through its Facebook and Twitter accounts, with the goal of enhancing political accountability.[24] Second, all possible means were used to ensure that the campaign remained non-partisan, including using the full set of municipalities for which audit reports pertaining to the incumbent mayor's government were available, randomizing treatment assignment within blocks of municipalities governed by the same political party, and avoiding the use of color schemes associated with any particular political party. Third, the information provided by Borde Político was publicly accessible and credibly measured by an independent federal audit institution, did not ask citizens to respond in any way to the information, and was unlikely to scramble voters' capacity to vote for their preferred candidate because prior studies had shown the use of FISM funds to be relevant to voters even during campaigns less focused on corruption (Arias et al. forthcoming; Chong et al. 2015; Larreguy, Marshall and Snyder 2020). We thus expected that the information would help voters make better-informed decisions. Finally, although high saturation campaigns inevitably concentrate information dissemination in ways that increase the possibility of affecting election outcomes, it is important to rigorously evaluate the effect of truthful information campaigns conducted by non-partisan civil society organizations on social media.

4 Exposure to Facebook ads

Over the course of Borde Político's accountability campaign, the Facebook ads appeared 7.3 million times on the screens of 2.7 million different Facebook users. More than 90% of engagements occurred via a smartphone, with the remainder primarily coming from desktop connections. We further analyze Facebook's analytics data for each ad campaign; in our case, a campaign targeted

[24] Borde Político had previously disseminated nationwide results of ASF reports online, and suggested delivering the information via Facebook to gauge how the effectiveness of their non-partisan campaigns could be maximized.

all precincts and segments assigned to receive a particular type of ad received in a municipality, and thus covered potentially multiple segments.

Aggregating the campaigns with and without the common knowledge slide by municipality, we examine engagement with the ads on Facebook by estimating OLS regressions of the form:

$$Y_m = \beta_1 Low\ saturation_m + \beta_2 High\ saturation_m + \mu_b + \varepsilon_m, \quad (4)$$

where Y_m is a measure of on-Facebook engagement per adult in the municipality where the ad was placed, and $Low\ saturation_m$ and $High\ saturation_m$ respectively indicate the low and high saturation municipal treatment conditions. The reference category is control municipalities, where outcomes are always zero because no ads were placed.

The results in panel A of Table 1 first demonstrate that the Facebook ad campaign ultimately reached a substantial fraction of targeted adults over its weeklong duration.[25] Column (1) shows that Borde Político's ad appeared on users' "News Feed" as paid content slightly more than once per targeted voter—that is to say, the number of adults divided by the municipal saturation level. Facebook's proprietary ad allocation mechanism thus ensured that paid-for ads appeared 0.32 times per adult member of the population in low saturation municipalities, and 0.95 times per adult in high saturation municipalities. Column (2) reports that a further 0.03 and 0.04 impressions per adult, or an additional 5-10% of views, came from organic views in low and high saturation municipalities. Organic views arise when friends on Facebook encounter Borde Político's ad because Facebook ad viewers shared, commented, or reacted to (e.g. liked) the ad. Turning to *unique* Facebook users in columns (3), the campaign reached more than one third of its intended population in the average municipality. Column (4) again indicates that user engagement contributed to a 5-10% increase in organic views. Although we cannot establish the intersection of respondents reached through paid-for ads and organic views or the location of organic views, these numbers suggest that at least a third of the adults in targeted segments of treated municipalities encountered the ad around three times on average.[26]

Beyond exposure, a non-trivial share of Facebook users actively engaged with the ad. Column (5) of panel A shows that approximately 2-3% of targeted voting-age adults publicly engaged with the ad (by sharing, commenting on, or liking the ad) or clicked through to the Facebook page. Furthermore, columns (7) and (9) respectively imply that 19% of targeted adults viewed the ads for at least 3 seconds and 11% of targeted adults viewed that ad for at least 10 seconds in low saturation municipalities, while the corresponding shares were 14% and 8% in high saturation municipalities.

[25] Appendix Figure A3 plots trends in Facebook ad engagement by day, indicating that the reach of the ad campaigns increased over the course of the week.

[26] Appendix Figure A4 displays the distribution of per-capita engagement by municipality, showing that engagement was generally greater in high saturation than in low saturation municipalities.

Table 1: Effect of municipal treatments on municipal Facebook ad engagement

	Municipal counts per capita (normalized by 2015 adult population)										
	Paid-for impressions (1)	Organic impressions (2)	Paid-for unique viewers (3)	Organic unique viewers (4)	Unique user page engagements (5)	Total views (of 3 seconds) (6)	Unique views (of 3 seconds) (7)	Total views (of 10 seconds) (8)	Unique views (of 10 seconds) (9)	Total views (of entire video) (10)	Unique views (of entire video) (11)
Panel A: average treatment effects											
High saturation	0.949***	0.042***	0.279***	0.024***	0.013***	0.181***	0.113***	0.075***	0.060***	0.043***	0.039***
	(0.103)	(0.007)	(0.026)	(0.004)	(0.002)	(0.018)	(0.011)	(0.007)	(0.006)	(0.004)	(0.004)
Low saturation	0.323***	0.030***	0.084***	0.016***	0.007***	0.064***	0.037***	0.028***	0.021***	0.017***	0.015***
	(0.080)	(0.009)	(0.019)	(0.005)	(0.002)	(0.014)	(0.008)	(0.006)	(0.005)	(0.004)	(0.003)
Observations	124	124	124	124	124	124	124	124	124	124	124
R^2	0.64	0.54	0.71	0.54	0.54	0.68	0.69	0.67	0.68	0.65	0.66
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
Control outcome std. dev.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Test: Low saturation = High saturation (p value, 2-sided)	0.000	0.157	0.000	0.114	0.002	0.000	0.000	0.000	0.000	0.000	0.000
Test: 4 * Low saturation = High saturation (p value, 2-sided)	0.258	0.019	0.420	0.022	0.050	0.159	0.238	0.124	0.178	0.067	0.084
Panel B: heterogeneity by municipal irregularities above and below the sample median											
High saturation	0.681***	0.037***	0.211***	0.019***	0.013***	0.129***	0.083***	0.059***	0.048***	0.036***	0.032***
	(0.118)	(0.012)	(0.032)	(0.007)	(0.003)	(0.021)	(0.013)	(0.010)	(0.008)	(0.007)	(0.006)
High saturation × Q3 or Q4	0.584**	0.018	0.143**	0.014	0.001	0.110**	0.063**	0.034*	0.027**	0.016	0.015*
	(0.235)	(0.017)	(0.059)	(0.010)	(0.004)	(0.042)	(0.025)	(0.017)	(0.014)	(0.010)	(0.009)
Low saturation	0.376***	0.026**	0.093***	0.014**	0.007**	0.064***	0.038***	0.030***	0.023***	0.019***	0.016***
	(0.135)	(0.012)	(0.030)	(0.006)	(0.003)	(0.021)	(0.013)	(0.010)	(0.008)	(0.007)	(0.006)
Low saturation × Q3 or Q4	-0.127	0.011	-0.026	0.006	-0.001	-0.006	-0.004	-0.007	-0.006	-0.003	-0.002
	(0.196)	(0.017)	(0.047)	(0.009)	(0.004)	(0.033)	(0.020)	(0.015)	(0.011)	(0.009)	(0.008)
Observations	124	124	124	124	124	124	124	124	124	124	124
R^2	0.70	0.59	0.75	0.59	0.56	0.72	0.73	0.70	0.71	0.67	0.68
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
Control outcome std. dev.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel C: heterogeneity by municipal irregularities quartile											
High saturation	0.703***	0.037**	0.216***	0.019**	0.013***	0.132***	0.085***	0.060***	0.048***	0.036***	0.032***
	(0.119)	(0.013)	(0.032)	(0.007)	(0.003)	(0.022)	(0.014)	(0.011)	(0.008)	(0.007)	(0.006)
High saturation × Q3	0.180	0.024	0.036	0.016	0.003	0.046	0.025	0.017	0.013	0.011	0.011
	(0.279)	(0.025)	(0.068)	(0.014)	(0.006)	(0.050)	(0.030)	(0.023)	(0.018)	(0.015)	(0.013)
High saturation × Q4	0.937**	0.012	0.240**	0.011	-0.000	0.166**	0.097**	0.050*	0.041*	0.021	0.020*
	(0.312)	(0.021)	(0.074)	(0.012)	(0.005)	(0.055)	(0.032)	(0.020)	(0.016)	(0.011)	(0.010)
Low saturation	0.383**	0.026*	0.096**	0.014*	0.007*	0.065**	0.039**	0.031**	0.024**	0.019**	0.017**
	(0.133)	(0.012)	(0.030)	(0.006)	(0.003)	(0.021)	(0.013)	(0.010)	(0.008)	(0.007)	(0.006)
Low saturation × Q3	-0.131	0.005	-0.032	0.002	-0.003	-0.008	-0.007	-0.011	-0.009	-0.006	-0.005
	(0.222)	(0.022)	(0.058)	(0.012)	(0.005)	(0.041)	(0.024)	(0.019)	(0.014)	(0.011)	(0.010)
Low saturation × Q4	-0.103	0.016	-0.017	0.010	0.000	-0.001	-0.000	-0.003	-0.003	0.000	0.000
	(0.225)	(0.021)	(0.052)	(0.012)	(0.005)	(0.039)	(0.023)	(0.016)	(0.012)	(0.011)	(0.009)
Observations	124	124	124	124	124	124	124	124	124	124	124
R^2	0.73	0.59	0.77	0.59	0.56	0.74	0.75	0.71	0.72	0.68	0.68
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
Control outcome std. dev.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes: User page engagements aggregates user decisions to share, comment, like, or click on the ad on Facebook. Each specification is estimated using OLS, and includes randomization block fixed effects. The two rump municipalities and the two municipalities for which electoral data is unavailable are excluded. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Column (11) indicates that the share of unique targeted adults that watched the entire ad was 8% and 5% in low and high saturation municipalities, respectively. The share of targeted adults that watched for at least 10 seconds is the best available approximation to the share of citizens exposed to the information about the fraction of spending subject to irregularities, which was reported 17 seconds into the ad. Nevertheless, Facebook users who did not get that far could still have clicked through to access a Facebook page that showed the level of irregularities or responded to the ad by thinking more on the issue or by discussing it with others. An average of around 15% of targeted voting-age adults (or around 20% of targeted Facebook users) thus substantively engaged with the ads, by watching at least 3 seconds of an ad, over the duration of Borde Político's Facebook ad campaign.

Although Facebook's algorithm dictated when ads appeared to Facebook users, the intended 1:4 ratio of exposure to ads across low and high saturation treatments was broadly maintained. The p values associated with the first test at the foot of panel A demonstrate that levels of engagement were systematically greater across all non-organic metrics of engagement in high saturation municipalities relative to low saturation municipalities. While the exposure ratio worked out to be around 1:3 in practice, the second test shows that we cannot statistically reject a 1:4 ratio for most measures of engagement.

The extent of Facebook user engagement with the ads did not generally vary with the level of irregularities reported in the ad. Interacting the saturation variables in equation (4) with our measures of irregularities, panels B and C of Table 1 show that—with the exception of high saturation municipalities with the highest amount of irregularities—citizen engagement did not significantly vary with the level of irregularities reported. For most comparisons, this suggests that differential treatment effects across quartiles of the irregularities distribution are unlikely to be driven by differential exposure or attention to the Facebook ads. Due to greater exposure, a given effect in municipalities where most irregularities were reported rests on lower persuasion rates.

Facebook's analytics data can also illuminate the *type* of users who were exposed to Borde Político's ads. Appendix Table A2 reports that the ads reached and were watched by men and women in equal proportion. Moreover, while the ads generally reached younger adults at relatively higher rates, the share of Facebook users that watched the ad was broadly in line with the 2010 Census age distribution. Ad consumption was greatest in the evening (37% of unique views), but also common in the morning (27%) and afternoon (31%). We do not observe systematic differences in these user characteristics by municipal saturation, irregularities quartile, or their interaction, suggesting that any differences in treatment effect by saturation level and irregularities quartile are unlikely to reflect differences in the types of voters that were exposed to the ads.

5 Effects of Facebook ads on voting behavior

We now evaluate the impact of the intervention on voting behavior. We first show that access to Borde Político's large-scale Facebook ads campaign slightly increased the incumbent party's vote share and turnout *on average*. Accounting for the level of irregularities reported, we then show that the information campaign substantially increased the vote share of incumbent parties whose mayors oversaw zero or negligible irregularities. While previous studies have similarly found that informed voters are more likely to reward the best-performing incumbent parties, the novelty of our finding is that non-partisan ads delivered via social media can produce effects that are at least as large. Moreover, we further establish that the large magnitude of the social media effect is principally driven by information campaign saturation: increases in support for the best-performing incumbent parties were significantly greater in the treated segments within high saturation than within low saturation municipalities. We then report within-municipality indirect effects that are almost as large in magnitude, before providing evidence as part of our robustness checks that show that indirect and differential saturation effects are not driven by mistargeted ads.

5.1 Average treatment effects

Column (1) of Table 2 reports our estimates of the *average* intent to treat effect of direct targeting by a Facebook ad campaign on municipal incumbent party vote share in targeted segments. Pooling across saturation and irregularities levels, we find that direct targeting of the Facebook ads treatment increased the incumbent party's vote share by 2.6 percentage points ($p = 0.10$) in treated segments. This 0.2 standard deviation increase represents around a 9% increase in incumbent support, relative to the 28% of registered voters that voted for the incumbent party in pure control segments in entirely untreated municipalities. This result, while not reaching conventional levels of statistical significance, is consistent with voters in the many municipalities where 0% (or negligible) irregularities were reported updating favorably about, or seeking to coordinate around, the incumbent party, or both. The next subsection shows that the average effect pools across differential effects by the level of reported irregularities.

The Facebook ad campaign also slightly increased electoral turnout on average. Appendix Table A16 reports that segments directly targeted with ads experienced a 1.3 percentage point increase in turnout ($p = 0.10$). Although the informational content of the ad differed significantly across municipalities, this increase in turnout is comparable to the positive effects of Facebook's own turnout mobilization campaigns in the US (see Bond et al. 2012; Jones et al. 2017). Since changes in turnout are small relative to changes in vote choice and the 64% turnout rate in the control group, and given that we do not detect significant heterogeneity in effects on turnout by the level of reported irregularities (see Appendix Table A16), we focus on incumbent party vote share

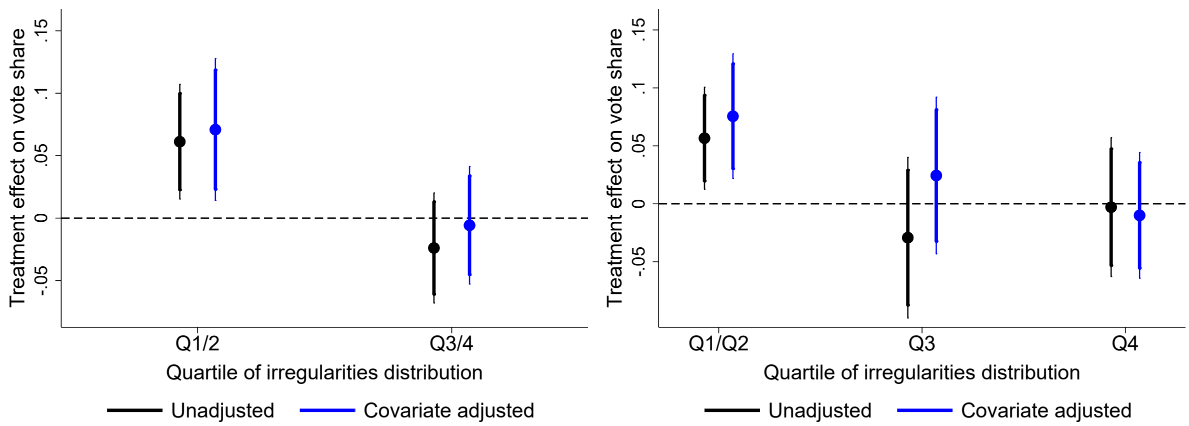
Table 2: Effect of Facebook ads on precinct-level municipal incumbent party vote share

	Incumbent party vote (share of turnout)				
	(1)	(2)	(3)	(4)	(5)
Facebook ads	0.026	0.061**	0.071**	0.057**	0.076***
	(0.016)	(0.023)	(0.029)	(0.022)	(0.027)
Facebook ads × Above-median irregularities		-0.085**	-0.077*		
		(0.034)	(0.042)		
Facebook ads × Irregularities Q3				-0.086*	-0.051
				(0.044)	(0.045)
Facebook ads × Irregularities Q4				-0.059	-0.085*
				(0.038)	(0.045)
Spillover	0.011	0.027	0.032	0.028	0.043*
	(0.016)	(0.022)	(0.026)	(0.021)	(0.025)
Spillover × Above-median irregularities		-0.031	-0.041		
		(0.036)	(0.041)		
Spillover × Irregularities Q3				0.001	-0.010
				(0.039)	(0.041)
Spillover × Irregularities Q4				-0.046	-0.065
				(0.044)	(0.046)
Observations	13,254	13,254	13,254	13,254	13,254
R^2	0.51	0.53	0.60	0.56	0.63
Control outcome mean	0.28	0.28	0.28	0.28	0.28
Control outcome std. dev.	0.14	0.14	0.14	0.14	0.14
Test: null effect of Facebook ads below median (p value)		0.01	0.02	0.01	0.01
Test: null effect of Facebook ads above median (p value)		0.29	0.81		
Test: null effect of Facebook ads in Q3 (p value)				0.41	0.48
Test: null effect of Facebook ads in Q4 (p value)				0.93	0.72
Test: null effect of spillover below median (p value)		0.23	0.23	0.18	0.09
Test: null effect of spillover above median (p value)		0.86	0.71		
Test: null effect of spillover in Q3 (p value)				0.35	0.26
Test: null effect of spillover in Q4 (p value)				0.64	0.50
Treatment × covariate interactions			✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Specifications including interactive covariates further include interactions between each treatment condition and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2 in columns (2)-(5). All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. The test statistics at the foot of the table report p values from two-sided hypothesis tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Figure 4: Conditional average treatment effect of Facebook ads on incumbent party vote share

(a) Below and above median irregularities (b) By quartile of the irregularities distribution



Notes: The estimates in Figure 4a derive from columns (2) and (3) of Table 2. The estimates in Figure 4b derive from columns (4) and (5) of the same table. Thick and thin lines denote 90% and 95% confidence intervals, respectively.

throughout this article.

5.2 Conditional average treatments effects by reported irregularities

Turning to the first of our main results, we next show that the relatively small average effects mask substantial heterogeneity with respect to the content of the information provided. First, the binary operationalization of malfeasance in column (2) of Table 2 shows that incumbent party vote share was 6.1 percentage points greater in segments targeted directly with Facebook ads in below-median municipalities, where zero or negligible irregularities were reported. This large and statistically significant effect represents a 20% increase, or a 0.44 standard deviation increase, relative to control group support for the incumbent party. In contrast, direct targeting reduced the incumbent party's vote share by 2.4 percentage points in municipalities where irregularities ranked above the median. While the interaction coefficient demonstrates that incumbent support was significantly lower in above-median than below-median municipalities, the test at the foot of column (2) indicates that the decline in vote share in above-median municipalities is not statistically significant ($p = 0.29$). Figure 4a depicts these conditional average treatment effects graphically.

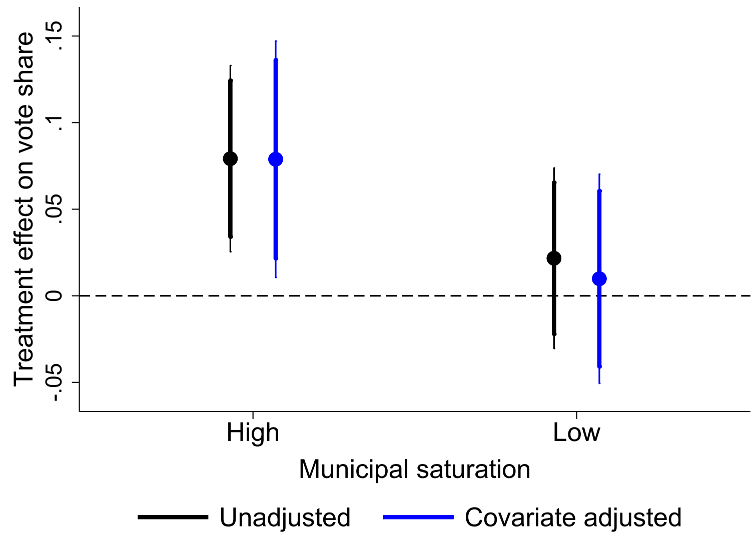
Second, the results are robust to dividing the distribution of municipal irregularities into quartiles. Recalling that quartiles Q1 and Q2 are pooled because 48% of municipalities engaged in zero irregularities, the first coefficient in column (4) again shows that Facebook ads directly targeting a segment increased the incumbent party's vote share by around 6 percentage points in municipalities where the government was responsible for zero or negligible irregularities. The negative interaction coefficients further indicate that the conditional average treatment effect is—as expected—smaller for higher levels of reported irregularities. This difference with quartiles Q1 and Q2 is statistically significant in Q3 of the irregularities distribution, but is not quite statistically significant at conventional levels in Q4 ($p = 0.11$). Figure 4b, together with the tests at the foot of column (4), show that the conditional average treatment effects in Q3 and Q4 (i.e. the sum of Q1/2 and respectively Q3 and Q4) cannot be distinguished from zero, again suggesting that only the electoral prospects of the municipal incumbent parties with the cleanest spending records were significantly affected by the Facebook ads.

The larger electoral effect of revealing good performance is in line with Arias et al.'s (forthcoming) prior findings in a smaller set of Mexican municipalities in 2015. The limited sanctioning of above-median and especially irregularities in Q4 could reflect a countervailing reduction in uncertainty about the incumbent's type among risk-averse voters (Arias et al. forthcoming), the possibility that bad performance is more likely to have already been reported in the media, or the somewhat lower levels of irregularities revealed in 2018 relative to prior years. Another more context-specific possibility is a lack of scope to reduce the incumbent party's vote share, given the substantial increase in support for MORENA in 2018 and the fact that MORENA was the incumbent in only one treated municipality. A floor of support, where only core supporters support malfeasant incumbent parties, could thus have been reached. This relates to Shadmehr and Bernhardt's (2017) prediction that the benefits of coordinating around signals of strong incumbent performance are greater where voters are already predisposed toward the challenger.

The magnitude of the conditional average effects—in particular the 6 percentage point increase in incumbent party vote share in treated segments within below-median municipalities—is similar to studies estimating the effects of relatively widespread media reporting of similar types of malfeasance in Brazil (e.g. Ferraz and Finan 2008). Moreover, our estimates are similar, but a bit larger than, the effects of medium saturation interventions (Arias et al. forthcoming; Banerjee et al. 2011; George, Gupta and Neggers 2019). In contrast, many experimental and quasi-experimental studies involve far lower levels of saturation and typically report effects of a couple of percentage points, as we systematically document in Appendix A.1. Given that the total cost of purchasing ads for Borde Político's non-partisan campaign was US$17,423, a one percentage point increase in votes for the best-performing incumbent parties within the average segment cost approximately US$11.[27] Illustrating the importance of mass dissemination of information, we next show that the large effects of non-partisan ads on Facebook are driven by the high saturation variant of the campaign.

[27] We compute the average campaign spending per segment (the total campaign cost divided by the number of treated segments), and then divide this by the treatment effect in Q1/2: $\frac{\$17422.57}{263}\frac{1}{100\times 0.06} = \$11.04$. However, this cost does not reflect the collection of ASF data, the production and targeting of the ads, and the credibility of Borde Político established through their prior work.

Figure 5: Conditional average treatment effect of Facebook ads on incumbent party vote share in municipalities with below-median irregularities, by ad campaign saturation



Notes: The estimates are from columns (2) and (3) of Table 3. Similar results hold when considering columns (4) and (5) of the same table. Thick and thin lines denote 90% and 95% confidence intervals, respectively.

5.3 Differential effects by municipal campaign saturation

Having established that voters targeted by Borde Político's accountability campaign rewarded law-abiding municipal incumbent parties, we next examine the extent to which ad campaign saturation amplified voters' electoral response to information provision. We accordingly estimate equation (2), which leverages the random assignment of treated municipalities to receive 20% or 80% information campaign saturation levels. In practice, exposure to Facebook ads was about three times greater in high than low saturation municipalities, as previously documented in Table 1.

The results in Table 3 demonstrate that greater levels of ad saturation indeed amplified the effects of Facebook ads reporting zero or negligible irregularities. Considering the binary measure of malfeasance, the estimates in column (2) indicate that revealing below-median levels of irregularities significantly increased the incumbent party's vote share in directly targeted segments by 7.9 percentage points—or around half a standard deviation—in high saturation municipalities. In contrast, the fifth row of the table shows that the analogous effect was 2.2 percentage points in directly treated segments within low saturation municipalities, and is statistically indistinguishable from zero. These differential effects by saturation in below-median municipalities are depicted in Figure 5. The test at the foot of column (2) shows that the 5.7 percentage point difference between the effects of targeting ads toward directly treated segments in high versus low saturation municipalities is statistically significant.

Column (4) of Table 3 reports similar results when municipalities are divided by quartile of

Table 3: Effect of Facebook ads on precinct-level municipal incumbent party vote share, by information campaign saturation

	Incumbent party vote (share of turnout)				
	(1)	(2)	(3)	(4)	(5)
Facebook ads in high saturation	0.031*	0.079***	0.079**	0.073***	0.080**
	(0.018)	(0.027)	(0.035)	(0.026)	(0.034)
Facebook ads in high saturation × Above-median irregularities		-0.120***	-0.072		
		(0.040)	(0.056)		
Facebook ads in high saturation × Irregularities Q3				-0.135**	-0.072
				(0.052)	(0.058)
Facebook ads in high saturation × Irregularities Q4				-0.077*	-0.042
				(0.044)	(0.062)
Facebook ads in low saturation	0.009	0.022	0.010	0.024	0.029
	(0.018)	(0.027)	(0.031)	(0.026)	(0.031)
Facebook ads in low saturation × Above-median irregularities		-0.024	-0.030		
		(0.041)	(0.047)		
Facebook ads in low saturation × Irregularities Q3				0.003	-0.028
				(0.044)	(0.052)
Facebook ads in low saturation × Irregularities Q4				-0.038	-0.051
				(0.050)	(0.056)
Spillover in high saturation	0.029	0.078***	0.071**	0.072***	0.072**
	(0.019)	(0.028)	(0.036)	(0.026)	(0.035)
Spillover in high saturation × Above-median irregularities		-0.124***	-0.065		
		(0.041)	(0.058)		
Spillover in high saturation × Irregularities Q3				-0.141**	-0.068
				(0.055)	(0.060)
Spillover in high saturation × Irregularities Q4				-0.079*	-0.029
				(0.043)	(0.064)
Spillover in low saturation	0.007	0.009	-0.003	0.011	0.017
	(0.019)	(0.026)	(0.031)	(0.026)	(0.031)
Spillover in low saturation × Above-median irregularities		0.002	0.002		
		(0.041)	(0.048)		
Spillover in low saturation × Irregularities Q3				0.050	0.023
				(0.044)	(0.053)
Spillover in low saturation × Irregularities Q4				-0.028	-0.035
				(0.050)	(0.056)
Observations	13,254	13,254	13,254	13,254	13,254
R^2	0.53	0.57	0.66	0.59	0.68
Control outcome mean	0.28	0.28	0.28	0.28	0.28
Control outcome std. dev.	0.14	0.14	0.14	0.14	0.14
Test: same effect of Facebook ads in high and low (p value)	0.27				
Test: same effect of Facebook ads in high and low below median (p value)		0.03	0.02	0.06	0.05
Test: same effect of Facebook ads in high and low above median (p value)		0.01	0.42		
Test: same effect of Facebook ads in high and low in Q3 (p value)				0.03	0.90
Test: same effect of Facebook ads in high and low in Q4 (p value)				0.78	0.24
Test: same effect of spillovers in high and low (p value)	0.30				
Test: same effect of spillovers in high and low below median (p value)		0.01	0.01	0.02	0.04
Test: same effect of spillovers in high and low above median (p value)		0.00	0.21		
Test: same effect of spillovers in high and low in Q3 (p value)				0.02	0.75
Test: same effect of spillovers in high and low in Q4 (p value)				0.64	0.20
Treatment × covariate interactions			✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Specifications including interactive covariates further include interactions between each treatment condition and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2 in columns (2)-(5). All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. The test statistics at the foot of the table report p values from two-sided hypothesis tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

the irregularities distribution. As the tests at the foot of the column attest, the best-performing incumbent parties were rewarded more in treated segments within high saturation municipalities than in treated segments within low saturation municipalities. Consistent with the lack of an effect in above-median municipalities and municipalities in Q3 and Q4 of the irregularities distribution, these tests also fail to detect robust differential effects across saturation levels within Q3 and Q4.

These results show that increasing the municipal-level saturation of the broader Facebook ad campaign substantially increased its effect on the average voter targeted *directly* by the campaign. This complementarity provides compelling evidence that saturation plays a causal role in magnifying the effects of information dissemination, and thus helps to explain why information campaigns are more likely to influence the voting behavior of treated voters when they are delivered by mass broadcast media (e.g. Ferraz and Finan 2008; Larreguy, Marshall and Snyder 2020). While other functions of media such as framing and credibility may also matter, our findings demonstrate that saturation is central to the impact of information distributed via digital communication technologies.

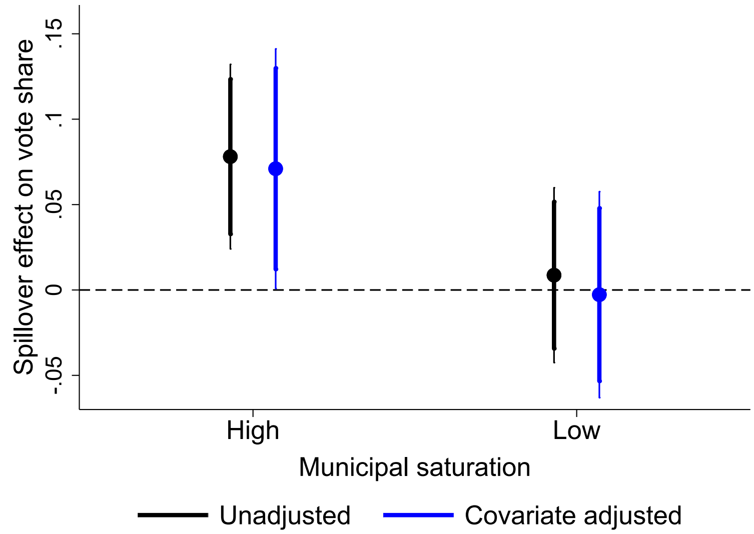
5.4 Spillover effects

We next assess spillover effects of the Facebook ads. Voters that were not directly targeted with ads may respond to the intervention because of online or offline interactions between voters, or due to party or media amplification of the ad's content. Such dynamics could occur within and across municipalities.

Turning first to spillover effects *within treated municipalities*, we leverage our two-level randomization to estimate cross-segment indirect effects by comparing untreated segments within both high and low saturation municipalities with segments in control municipalities. Plotting the indirect effect estimates for municipalities with irregularities below the median level, Figure 6 shows that untreated segments in high saturation municipalities were about 7 percentage points more likely to vote for the best-performing incumbent parties. These indirect effects on voting behavior are almost as large as in targeted segments, and not statistically distinguishable in magnitude. In contrast, the indirect effect of around 1 percentage point in low saturation municipalities is not distinguishable from zero; the test at the foot of column (2) of Table 3 demonstrates that the indirect effect is significantly larger in high relative to low saturation municipalities. Again, while largely negative, we cannot detect significant indirect effects of the ads in segments in municipalities where larger levels of irregularities were reported. We explore the propagation mechanisms that high saturation campaigns may have activated to produce these large spillover effects where the incumbent party performed best in Section 6.

It is also possible that the effects of Facebook ads could extend *across municipalities*. While our experiment was designed to estimate indirect effects within treated municipalities, we did not pre-

Figure 6: Conditional average spillover effect of Facebook ads on incumbent party vote share in municipalities with below-median irregularities, by ad campaign saturation



Notes: The estimates are from columns (2) and (3) of Table 3. Similar results hold when considering columns (4) and (5) of the same table. Thick and thin lines denote 90% and 95% confidence intervals, respectively.

specify an approach to estimate spillover effects across municipalities. We explore this possibility by examining 640 electoral precincts in municipalities that are not in the experimental sample, but also held elections in 2018 and are located within 5 kilometers of a single segment in an experimental municipality.[28] This enables us to exploit the same experimental variation to estimate across-municipality effects using equations (1) and (2).[29] To capture effects on the party about which information is revealed in the nearby experimental municipality, the outcome variable is the vote share of the party (or coalition) that included the party that governs the nearby municipality in our experimental sample.

Although these tests leverage only around 30% of the segments used for the main analysis, the results reported in Appendix Table A17 suggest that the effects of the Facebook ad campaigns do not extend beyond the municipalities where the ads were disseminated. Relative to control segments in untreated municipalities, column (1) reports no evidence to suggest that segments that were directly or indirectly targeted with Facebook ads significantly affected the vote choices of voters in nearby precincts within other municipalities on average. Columns (2)-(5) further detect no systematic evidence of conditional effects analogous to those in treated municipalities. These results suggest that the impact of Facebook ads does not spill across municipalities; this also im-

[28]We exclude electoral precincts within 5 kilometers of multiple experimental segments in order to cleanly define the cross-municipality spillover treatment.

[29]We exclude the block fixed effects, since not all experimental municipalities are nearby to non-experimental precincts in this cross-municipality spillover sample.

plies that comparisons within the experimental sample are unlikely to be contaminated by cross-municipality interference.

5.5 Robustness checks

We probe the robustness of our findings by exploring statistical significance through randomization inference, by checking for outliers by dropping each randomization block, by testing for Facebook ad mistargeting, and by interactively adjusting for potential correlates of municipal irregularities.

5.5.1 Randomization inference

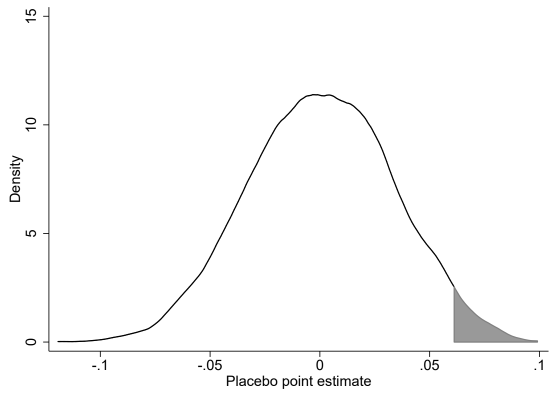
Treatment conditions are well-balanced across observable predetermined characteristics both above and below the median level of irregularities and within each quartile, and all estimates adjust for prior incumbent party vote share. Nevertheless, it remains possible that a fluke random assignment could account for the large effects of Facebook ads on the vote shares of the best-performing incumbent parties. We probe this possibility by replicating the exact randomization procedure to generate 5,000 alternative treatment assignment vectors. Focusing on our main findings in municipalities where irregularities were below the median level, only 2.9% and 3.3% of alternative randomizations yield a larger positive direct effect of Facebook ads when pooling across municipal saturation levels in the by-median and by-quartile specifications, respectively. Analogous tests show that only 1.6% and 2.0% of alternative randomizations yield a larger positive direct effect of Facebook ads in high saturation municipalities, and only 6.2% and 9.2% of alternative randomizations yield a larger differential effect of Facebook ads in high relative to low saturation municipalities. For the indirect effects in municipalities below the median level of irregularities, which we discuss below, the corresponding shares of alternative randomizations that yield more positive effects are 22.7% and 20.7%, 1.7% and 2.2%, and 2.9% and 5.1%. Figure 7 shows the distribution of placebo coefficients for municipalities below the median level of irregularities. Randomization inference thus reinforces the robustness of our findings.

5.5.2 Removing potential outlier blocks

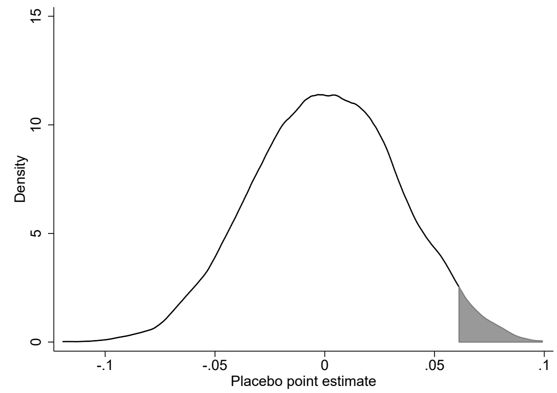
While the permutation tests just described demonstrate robustness to alternative treatment assignments, it remains possible that our findings could reflect large effects in outlier municipalities. We investigate this using a leave-one-out approach, where we separately re-estimate our main specifications without each of the 42 blocks in the sample. The results in Appendix Figure A6 show that the point estimates for the direct and indirect effects in below-median municipalities, as well as the differential saturation effects, remain stable after dropping any individual randomization block.

Figure 7: Observed treatment effects within the distribution of placebo treatment assignments

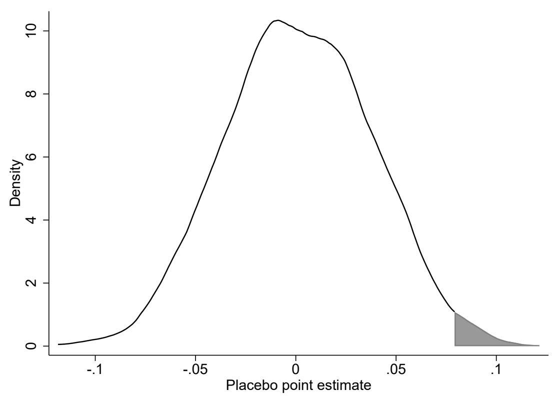
(a) Direct effect in municipalities with below median irregularities



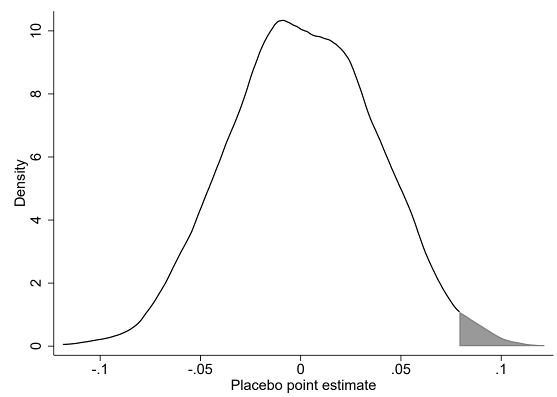
(b) Indirect effect in municipalities with below median irregularities



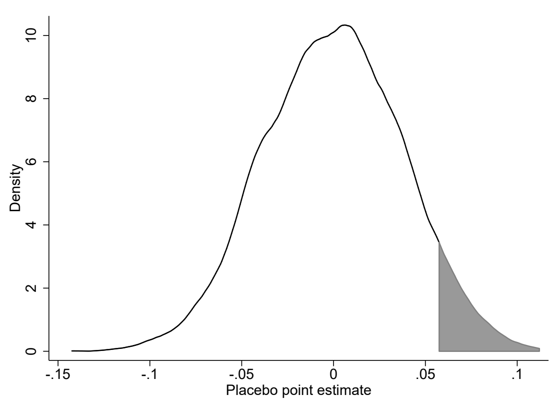
(c) Direct effect in high saturation municipalities with below median irregularities



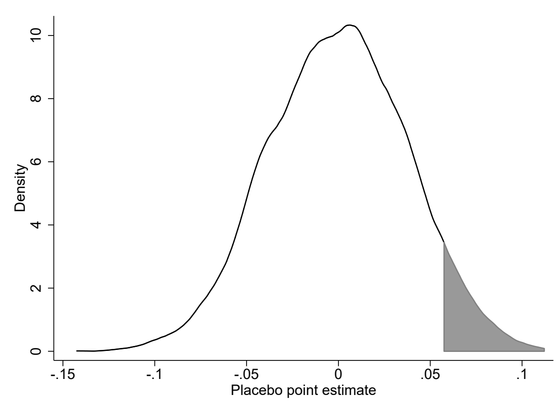
(d) Indirect effect in high saturation municipalities with below median irregularities



(e) Difference in direct effect between high and low saturation municipalities with below median irregularities



(f) Difference in indirect effect between high and low saturation municipalities with below median irregularities



Notes: Holding the observed outcomes fixed, each distribution reports the distribution of treatment effects based on 5,000 placebo replications of our randomization. The shaded area denotes permuted treatment assignments producing larger treatment effects than our observed treatment effect.

5.5.3 Potential mistargeting of Facebook ads

The integrity of the treatment conditions relies on Facebook targeting Borde Político's informational ads at users that reside—and thus vote—in treated segments. However, despite more than 90% of Facebook users encountering the ads on mobile devices, mistargeting could still arise for users without reliable GPS data or who commute across segment boundaries. Alternatively, mistargeting could arise where Facebook targets ads at users based on current location, rather than home address.

How such mistargeting would bias our estimates depends on the extent to which segments experience *inward mistargeting* (citizens within a given segment receiving ads intended for other segments in their municipality) and/or *outward mistargeting* (citizens in a given segment not receiving ads intended for their segment). In the presence of mistargeting, comparisons between directly treated and pure control segments are likely to capture a lower bound on the average or conditional average effect of perfectly targeted ads. This is because targeted segments could only experience net outward mistargeting that reduces the intensity of treatment, while mistargeting could not reduce—and is unlikely to affect—intended ad targeting in control segments in untreated municipalities. In contrast, comparisons between non-targeted segments within treated municipalities and control segments would overstate the indirect effect of Facebook ads because mistargeting would induce direct exposure to ads intended for the directly targeted segments within the same municipality. Moreover, differences between high and low saturation municipalities—for both direct and indirect effects—would be overstated in the presence of mistargeting because inward mistargeting would be greater in high saturation municipalities than in low saturation municipalities because more other segments within high saturation municipalities are directly treated.

A simple accounting exercise indicates that the amount of mistargeting required to account for our estimates is implausibly substantial. Let $\eta_s \in [0,1]$ denote the share of ads targeted at treated segments that reach other (treated or untreated) segments due to mistargeting within a 5-segment municipality with saturation level $s \in \{0.2, 0.8\}$, and let κ_s denote the probability that a targeted voter is exposed to a given ad. We then compute the probability of exposure to the Facebook ads in directly and indirectly targeted segments within municipalities treated with saturation level s as follows:[30]

$$
\begin{aligned}
\Pr[ad\ exposure|targeted,\ s=0.2] &= (1-\eta_{0.2})\kappa_{0.2}, & (5)\\
\Pr[ad\ exposure|not\ targeted,\ s=0.2] &= \frac{\eta_{0.2}}{4}\kappa_{0.2}, & (6)\\
\Pr[ad\ exposure|targeted,\ s=0.8] &= \left(1-\eta_{0.8}+\frac{3\eta_{0.8}}{4}\right)\kappa_{0.8}, & (7)\\
\Pr[ad\ exposure|not\ targeted,\ s=0.8] &= \eta_{0.8}\kappa_{0.8}. & (8)
\end{aligned}
$$

Assuming that the effects of treatment are proportional to the exposure rate, the degree of mistargeting required to explain our main results in Q1/2 can be computed by setting the exposure ratio between targeted and non-targeted segments at a given saturation level equal to the ratio of treatment effects in Q1/2 reported in column (4) of Table 2. For example, in low saturation municipalities, this implies: $\frac{1-\eta_{0.2}}{\frac{\eta_{0.2}}{4}} = \frac{0.024}{0.011}$. Solving this equation, and an analogous one for high saturation municipalities, yields $\eta_{0.2} = 0.65$ and $\eta_{0.8} = 0.79$. These estimates suggest that at least two thirds of ads would need to have been mistargeted to account for our findings.

Although Facebook does not provide the ad exposure data by individual or exact location needed to directly detect mistargeting, such levels of mistargeting are highly unlikely. Nevertheless, we also test two implications of mistargeting to further gauge whether substantial mistargeting could plausibly account for the indirect and differential saturation effects.

First, we address the possibility that Facebook targeted ads based on current location and thus mistargeted individuals that moved across segments. If this were the case, then indirect effects and the differential effect in high saturation municipalities should be greater in municipalities where citizens move around more regularly, e.g. commuting or purchasing goods from markets. We measure mobility, using pre-pandemic 2020 data from Facebook's Data For Good dataset, as the average number of level-16 Bing tiles (which are approximately 600 meters by 600 meters in area at the equator) visited by Facebook users in a 24-hour period.[31] The results in Appendix Table A18 provide no evidence to suggest that the magnitude of indirect effects and direct or indirect effects in high saturation municipalities were greater in municipalities where Facebook users move across tiles more often.

Second, we leverage random variation in the operation of ad campaigns that could generate systematic differences in exposure to ads in the event that Facebook mistargeted ads. In high sat-

[30]By taking expectations over segments, no assumptions are required about the spatial structure of segments. Similar, but slightly more complicated, formulas apply to the small number of municipalities with more than 5 segments.

[31]See https://research.fb.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response for more details about the mobility measure and www.docs.microsoft.com/en-us/bingmaps/articles/bing-maps-tile-system for more information about Bing tiles. We use this mobility measure for two reasons. First, we are confident it reflects structural patterns of mobility between connected geographies. Second, these fine-grained data come directly from Facebook and thus we presume they may be considered for its ad targeting algorithm.

uration municipalities (and low saturation municipalities containing 10 or more segments), two separate ad campaigns were conducted—with and without a common knowledge screen at the end of the video. For example, in a high saturation municipality with 5 segments, 2 segments were randomly assigned to receive the ads from the common knowledge (CK) ad campaign and 2 other segments were randomly assigned to receive ads from a separate ad campaign that did not include the common knowledge screen (NCK). While the minor difference in content between the CK and NCK campaigns did not affect voting behavior, as Appendix Table A13 demonstrates, the occurrence of two campaigns within the same municipality created the potential for differential exposure due to mistargeting. Because each type of campaign was set up separately on Facebook, individual exposure to a CK ad did not influence the probability of exposure to a NCK ad. Consequently, if mistargeting occurs at the edge of the areas targeted with Facebook ads, this implies that electoral precincts near a border between two adjacent segments assigned to receive *different* ad campaigns could be targeted by both campaigns. In contrast, no such double exposure would occur in electoral precincts near a border between two adjacent segments assigned to the *same* ad campaign.

Our second test for potential mistargeting thus compares voting behavior between electoral precincts near the border between two adjacent treated segments that were part of different ad campaigns (one CK and one NCK) with electoral precincts near the border between two adjacent treated segments that were part of the same ad campaign (either both CK or both NCK). Observing more pronounced effects in line with the results in Table 2 among precincts at the border of different campaigns would suggest mistargeting. We test this by leveraging the fact that, conditional on the number of treated segments that are near to a given precinct within a treated segment (within 100 meters of the precinct, but in a different segment), the share of such nearby segments assigned to the same vs. different ad campaigns is random.[32] Since Appendix Table A19 shows that incumbent party vote share is not increasing in the share of nearby treated segments that were part of different ad campaigns, we again find no evidence to suggest that mistargeting of this form drives the results.

Neither test can definitively rule out ad mistargeting. Nevertheless, passing each test—combined with Facebook's claim to precisely identify users' nighttime locations—suggests that mistargeting is unlikely to be the main driver of the large effects we observe. Moreover, it is highly unlikely to reach the high rates of mistargeting required to fully explain our estimates.

[32]To see why, consider an 80% saturation municipality that comprises 5 segments, of which 2 are targeted with the NCK campaign and 2 are targeted with the CK campaign. For any given precinct within a treated segment that is adjacent to 1 other treated segment, the share of adjacent treated segments from a different campaign is 0 (with probability $\frac{1}{3}$) or 1 (with probability $\frac{2}{3}$). For any given precinct within a treated segment that is adjacent to 2 other treated segments, the share of adjacent treated segments from a different campaign is 0.5 (with probability $\frac{1}{3}$) or 1 (with probability $\frac{2}{3}$). For municipalities with multiples of 5 segments, the probability distributions differ a little, but the logic is the same.

5.5.4 Adjusting for differences in malfeasance across municipalities

While exposure to Facebook ads and municipal saturation were randomly assigned, the share of irregularities reported was not. Consequently, heterogeneity in response to Facebook ads across municipalities where the ASF detected different levels of irregularities could instead reflect other differences across these municipalities.

We address this risk of confounding by adjusting for the interaction between treatment conditions and 11 predetermined municipal-level covariates that could plausibly moderate the effects of treatment. Our first set of interactive covariates include other municipality-specific quantitative information conveyed by the ad—the financial year to which the audit pertained and the (log) amount of FISM funds received per capita by the municipality in that year—to address the concern that ad content beyond the share of irregularities affected voting behavior. Second, we include prior municipal incumbent party vote share to allow audit report revelations to differentially impact the parties of more or less popular local governments. Finally, we include eight demographic and socioeconomic variables that could either be correlated with reported irregularities and facilitate coordinated responses to ads or could proxy for voters' prior beliefs, attentiveness or access to Facebook ads, or capacity to comprehend such ads.[33] Summary statistics for each covariate are reported in Appendix Table A8 by quartile.

The estimates in columns (3) and (5) of Tables 2 and 3 adjust for interactions between each treatment condition and all 11 covariates simultaneously. Although reducing the residual variation in irregularities across municipalities increases the standard errors, the covariate-adjusted estimates are largely similar to the unadjusted estimates in columns (2) and (4). We again find a large positive effect of Facebook ads in Q1/2 municipalities, which is driven by treated segments in high saturation municipalities. The conditional effects in Q3 and Q4 remain null.

6 Mechanisms

A natural starting point for interpreting the effects on voting behavior is that *viewers* of the ads on Facebook were persuaded to support municipal incumbent parties that engaged in zero/negligible expenditure irregularities. However, while some viewers may have been directly persuaded, this mechanism alone cannot plausibly fully account for the direct effect of targeting reported in Tables 2 and 3. If Borde Político's Facebook ad campaign only influenced citizens that viewed the ad on Facebook, either through paid-for or organic impressions, the implied persuasion rates far exceed

[33]These municipal-level covariates are: (log) 2010 adult population; average years of schooling in 2010; the share illiterate in 2010; the average number of occupants per room, by household, in 2010; the average number of children per woman in 2010; the share of households with electricity, water, and drainage in 2010; the share of the municipal population that is working age in 2010; and the share of households with internet at home in 2010.

prior studies (DellaVigna and Gentzkow 2010).[34] Moreover, given the absence of significant mistargeting, direct persuasion of Facebook ad viewers cannot explain the large spillover effects or the differential direct and indirect effects observed across municipal saturation levels. Consequently, as Guriev, Melnikov and Zhuravskaya (2021) also note, the "atomized viewer" assumption underpinning persuasion rate calculations—that the treatment only affected citizens exposed to the ad on Facebook—is unlikely to hold for mass ad campaigns on social media.

We explore two classes of channel through which citizens that did not view the ads on Facebook might also have changed their vote choice after ads were disseminated in their municipality, and more so when municipal saturation was higher. First, social interactions catalyzed by citizens that consumed the ad on Facebook may influence the voting behavior of other citizens. Second, local media outlets or political campaigns—which could also influence large numbers of citizens—may have produced related messaging or content in response to Borde Político's Facebook ad campaign. We assess each propagation mechanism in turn, using tests that move beyond our pre-analysis plan to show that the results are driven primarily by social interactions between citizens.

6.1 Information campaigns and social interactions between citizens

As Druckman, Levendusky and McLain (2018) have shown for partisan media in the US, media's substantial influence can belie a limited audience because its content propagates through social networks to affect individuals who were not exposed directly. First, interpersonal discussion—whether on or offline—could have diffused the content of Borde Político's Facebook ads to citizens that would not have otherwise received the information (Alt et al. 2022) or provided the basis for conversations that persuaded citizens that had not encountered the ads directly to alter or coordinate their vote choices (Druckman, Levendusky and McLain 2018; Katz and Lazarsfeld 1955). Such mechanisms are likely to be greatest within highly-connected social networks (Caprettini et al. 2021). Although we can only capture content sharing within Facebook, Table 1 shows that organic impressions accounted for around 10% of the individuals that were exposed to the ad on Facebook. The relatively high rates of exposure to Borde Político's campaign also make it likely that the ads reached opinion leaders and information disseminators within family and friend networks.

[34] The persuasion rate when we assume that the ads could only affect the 7% of citizens in directly targeted segments who reached the 17th second of the ad—when the share of irregularities appeared in the video—is 71%. Specifically, linear interpolation of the unique viewer estimates in columns (9) and (11) of Table 1 implies that $100 \times \frac{0.015+(0.021-0.015)\times\frac{26-17}{26-10}}{0.2} \approx 9.2\%$ of targeted voters in low saturation municipalities and $100 \times \frac{0.038+(0.060-0.038)\times\frac{26-17}{26-10}}{0.8} \approx 6.3\%$ of targeted voters in high saturation municipalities reached the 17th second of the 26-second ad. Given differences in targeting intensity across low and high saturation municipalities, this yields a sample average exposure rate of 6.9%. Following DellaVigna and Gentzkow (2010), the persuasion rate in Q1/2—derived from the estimate in column (2) of Table A15, the analog of Table 2 for incumbent vote share as a share of registered voters, the relevant denominator for computing persuasion rates—is then: $100 \times \frac{0.044}{0.069}\frac{1}{1-0.18} \approx 78\%$. Even assuming that watching only 3 seconds of the ad is sufficient to prime viewer responses, the implied persuasion rate is still 35%.

Second, communication induced by the ads could further generate common knowledge that others have received the information. This may, in turn, amplify the effects of information dissemination by inducing voters to tacitly coordinate their choices around public signals (Morris and Shin 2002)—especially where many citizens receive the signal (Cornand and Heinemann 2008)—or to make explicit agreements about how to vote (Shadmehr and Bernhardt 2017). Such mechanisms are plausible in Mexico, where 31% of Mexicans reported talking to neighbors, 44% to friends, and 56% to family about electoral campaigns sometimes or often in 2018.[35]

Since social networks extend across segments within municipalities, interactions that arise after some nodes in a network are exposed to Facebook ads could influence citizens in both directly and indirectly targeted segments within treated municipalities. Given the large direct and indirect effects already documented, we test two more distinctive implications of effects driven by social mechanisms.

6.1.1 Effects of Facebook ads increase with municipal social connectedness

We first probe the social network mechanism by examining whether the direct and indirect of Facebook ads are greater in more socially-connected municipalities. We proxy for network connectedness using Facebook's Social Connectedness Index (SCI); this calculates the probability that any two Facebook users within a given municipality are friends on Facebook.[36] In addition to increasing exposure to any politically relevant content produced by viewers on Facebook, friendship is a good proxy for communication between individuals off Facebook as well (e.g. Gilbert and Karahalios 2009). We thus expect the SCI to correlate with the types of social networks where information diffusion, interpersonal persuasion, or common knowledge are most likely to kick in.

The results in Table 4 confirm that the direct and indirect effects of Facebook ads increase with a municipality's social connectedness. Each coefficient reports how the conditional average treatment effects of Facebook ads by reported irregularities, at each municipal saturation level, varies with the (standardized) SCI index.[37] Among municipalities below the median level of irregularities, the estimates in columns (1)-(4) indicate that a standard deviation increase in the SCI is associated with at least a 6 percentage point greater positive effect of Facebook ads on the incumbent party's vote share. These significant heterogeneous effects hold across direct and indirect effects and across

[35] Based on the Comparative National Election Project's 2018 Mexico survey.

[36] The SCI is computed as $SCI_m = \frac{Friendships_m}{Users_m^2}$, where $Friendships_m$ is the number of Facebook friendships between Facebook users in municipality m, and $Users_m$ is the number of Facebook users in municipality m. See https://dataforgood.facebook.com/dfg/docs/methodology-social-connectedness-index for further information. We standardize the SCI, which is available for all but one municipality in our sample, to facilitate interpretation. Unfortunately, Facebook does not provide the microdata required to construct further measures of network structure of average centrality.

[37] By standardizing the SCI index using the weights used in our regressions, we focus on heterogeneity in the conditional average treatment effects by holding the average effect constant.

Table 4: Differential effect of Facebook ads on precinct-level municipal incumbent party vote share, by Facebook's municipal-level social connectedness index

	Incumbent party vote (share of turnout)			
	(1)	(2)	(3)	(4)
Facebook ads in high saturation × SCI (standardized)	0.068**	0.101*	0.070**	0.118**
	(0.034)	(0.056)	(0.035)	(0.055)
Facebook ads in high saturation × Above-median irregularities × SCI (standardized)	-0.077	-0.176**		
	(0.054)	(0.085)		
Facebook ads in high saturation × Irregularities Q3 × SCI (standardized)			-0.056	-0.069
			(0.045)	(0.073)
Facebook ads in high saturation × Irregularities Q4 × SCI (standardized)			-0.039	-0.267***
			(0.066)	(0.099)
Facebook ads in low saturation × SCI (standardized)	0.083*	0.061	0.081*	0.116*
	(0.047)	(0.070)	(0.047)	(0.059)
Facebook ads in low saturation × Above-median irregularities × SCI (standardized)	-0.066	-0.046		
	(0.053)	(0.061)		
Facebook ads in low saturation × Irregularities Q3 × SCI (standardized)			0.013	0.044
			(0.054)	(0.060)
Facebook ads in low saturation × Irregularities Q4 × SCI (standardized)			-0.122**	-0.152***
			(0.048)	(0.055)
Spillover in high saturation × SCI (standardized)	0.070**	0.111**	0.072**	0.128**
	(0.033)	(0.056)	(0.034)	(0.055)
Spillover in high saturation × Above-median irregularities × SCI (standardized)	-0.077	-0.141		
	(0.054)	(0.089)		
Spillover in high saturation × Irregularities Q3 × SCI (standardized)			-0.009	0.012
			(0.050)	(0.073)
Spillover in high saturation × Irregularities Q4 × SCI (standardized)			-0.087	-0.284***
			(0.059)	(0.101)
Spillover in low saturation × SCI (standardized)	0.102**	0.085	0.100**	0.146**
	(0.042)	(0.070)	(0.042)	(0.061)
Spillover in low saturation × Above-median irregularities × SCI (standardized)	-0.086	-0.062		
	(0.052)	(0.063)		
Spillover in low saturation × Irregularities Q3 × SCI (standardized)			-0.004	0.031
			(0.051)	(0.060)
Spillover in low saturation × Irregularities Q4 × SCI (standardized)			-0.150***	-0.177***
			(0.044)	(0.056)
Observations	13,233	13,233	13,233	13,233
R^2	0.62	0.72	0.65	0.75
Control outcome mean	0.28	0.28	0.28	0.28
Control outcome std. dev.	0.14	0.14	0.14	0.14
Treatment × covariate interactions		✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Lower-order interaction terms are included, but omitted to save space. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

low and high saturation municipalities, suggesting that social networks may partially substitute for higher levels of ad saturation. The difference in effect relative to municipalities with more irregularities is also greater in municipalities with a higher SCI, although the significance of the above-median, Q3, and Q4 terms is less consistent.

While direct persuasion of ad viewers may explain some changes in voting behavior, this evidence—together with the spillover and saturation effects reported previously—suggests that social interactions that occurred on or off of Facebook likely amplified the impact of incumbent performance information disseminated via Facebook ads. Social propagation thus appears to play a key role in driving the impact of online information campaigns.

6.1.2 More coordinated vote choices

Next, we examine whether part of the effect driven by social interactions reflects tacit or explicit coordination of vote choices by voters. In each electoral precinct, we compute the effective number of political parties by vote share (ENPV)—a common measure of vote concentration in political science (Laakso and Taagepera 1979), that is the inverse of the Herfindahl-Hirschman Index.[38] If Facebook ads, especially in high saturation municipalities, coordinate voting behavior around the best-performing incumbent parties and their main opponents, we should expect the ENPV to decline.

One potential concern is that changes in ENPV could reflect both strategic coordination of votes around the main contenders and sincere changes in support for the incumbent. However, we show in Appendix A.14 that an increase in support for a party—as we find in Q1/2 of the irregularities distribution—will instead increase the ENPV when that party is not too popular relative to the parties that it gains votes from. Since the average incumbent party in control municipalities only received 28% of votes, any change in ENPV due to a pure information effect is expected to bias *against* detecting more concentrated vote choices.

The results in Figure 8 indicate that treatment decreased the ENPV in precincts that were directly and indirectly targeted with Facebook ads when the ads targeted large shares of a municipality.[39] In high saturation municipalities with below-median levels of irregularities, precincts in directly and indirectly targeted segments experienced around a 0.3-0.4 reduction in effective parties—roughly, a 10% decline relative to the average ENPV in precincts in control segments. Again, negligible reductions in the ENPV are observed in similar low saturation municipalities. Moreover, Appendix Table A21 reports similar results at the segment level, suggesting that the level at which coordination occurs may be larger than the city block or rural locality that a precinct

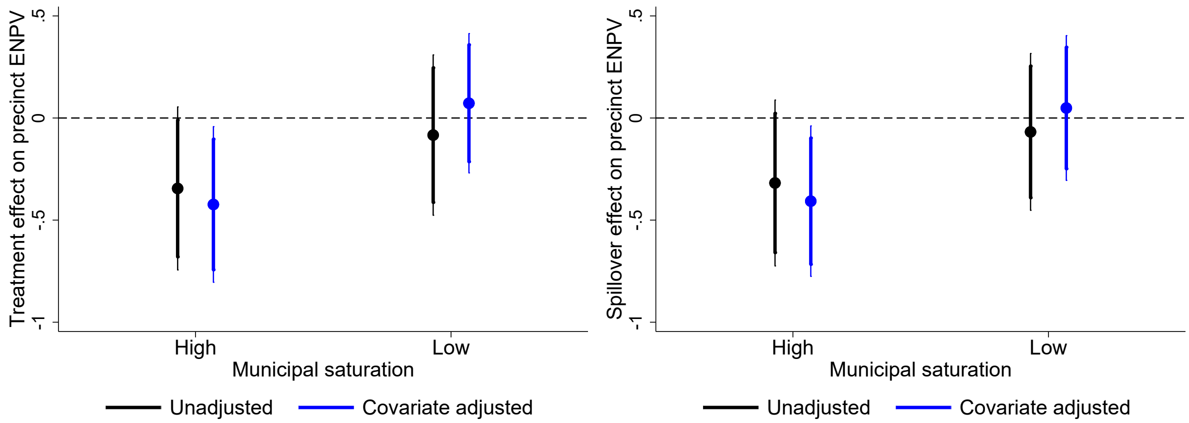
[38]The ENPV is defined as: $ENPV_{pm} = \frac{1}{\sum_{j=1}^{J} v_{jpm}^2}$, where v_{jpm} is the vote share of party $j = 1, ..., J$ in precinct p of municipality m.

[39]The underlying regression estimates are reported in Appendix Table A20.

Figure 8: Conditional average treatment effect of Facebook ads on precinct-level ENPV

(a) Direct effects in municipalities with below median irregularities

(b) Indirect effects in municipalities with below median irregularities



Notes: The estimates in each figure derive from columns (2) and (3) of Appendix Table A20. Thick and thin lines denote 90% and 95% confidence intervals, respectively.

typically comprises.

Consistent with the importance of social interactions, these findings suggest that part of the large impact of high saturation information campaigns may have been driven by vote coordination. Further parsing whether this occurred through direct discussion or common knowledge is beyond the scope of our study.

6.1.3 No spillover to support for MORENA in concurrent presidential elections

Another potential mechanism through which Facebook ads could have affected vote choices is by increasing the salience of malfeasance. Due to the party's anti-corruption messaging, voters regarding malfeasance as a key issue were likely to vote for MORENA. Since MORENA was the incumbent party in hardly any municipalities, the issue-priming mechanism produces equivalent predictions to learning or coordination mechanisms. However, increased concern about malfeasance is likely to spill up to the concurrent presidential election, where corruption was central to López Obrador's anti-establishment campaign. If Borde Político's Facebook ad campaign increased the salience of government malfeasance, the ads are likely to increase López Obrador's vote share, especially in municipalities where significant levels of irregularities were reported.

Table 5 reports effects of Facebook ads on López Obrador's presidential vote share by level of reported irregularities. Our estimates show that Facebook ads did not significantly affect support for MORENA, either on average or in municipalities where greater malfeasance was reported. The lack of effect at the presidential level suggests that the substantial effect of the ads on municipal election outcomes is unlikely to be driven by the ads increasing the salience of malfeasance in the

Table 5: Effect of Facebook ads on precinct-level vote share for López Obrador in the presidential election

	MORENA/PT/PES vote (share of turnout)				
	(1)	(2)	(3)	(4)	(5)
Facebook ads	-0.014	-0.023	-0.007	-0.020	-0.009
	(0.016)	(0.021)	(0.024)	(0.021)	(0.023)
Facebook ads × Above-median irregularities		0.036	-0.021		
		(0.035)	(0.037)		
Facebook ads × Irregularities Q3				0.011	-0.049
				(0.049)	(0.044)
Facebook ads × Irregularities Q4				0.053	0.002
				(0.041)	(0.043)
Spillover	-0.007	0.006	0.003	0.007	-0.002
	(0.016)	(0.022)	(0.024)	(0.023)	(0.024)
Spillover × Above-median irregularities		-0.025	-0.028		
		(0.039)	(0.037)		
Spillover × Irregularities Q3				-0.065	-0.041
				(0.049)	(0.047)
Spillover × Irregularities Q4				0.012	-0.015
				(0.044)	(0.040)
Observations	13,267	13,267	13,267	13,267	13,267
R^2	0.60	0.61	0.70	0.61	0.70
Control outcome mean	0.49	0.49	0.49	0.49	0.49
Control outcome std. dev.	0.18	0.18	0.18	0.18	0.18
Test: null effect of Facebook ads below median (*p* value)		0.27	0.76	0.33	0.70
Test: null effect of Facebook ads above median (*p* value)		0.64	0.19		
Test: null effect of Facebook ads in Q3 (*p* value)				0.83	0.08
Test: null effect of Facebook ads in Q4 (*p* value)				0.34	0.81
Test: null effect of spillover below median (*p* value)		0.78	0.91	0.76	0.95
Test: null effect of spillover above median (*p* value)		0.51	0.24		
Test: null effect of spillover in Q3 (*p* value)				0.14	0.21
Test: null effect of spillover in Q4 (*p* value)				0.60	0.53
Treatment × covariate interactions			✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable (MORENA coalition vote share in the 2015 legislative elections) and randomization block fixed effects. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. The test statistics at the foot of the table report p values from two-sided hypothesis tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

eyes of voters.

6.2 Limited political and media responses to the Facebook ad campaigns

The main alternative class of mechanisms that could account for the campaign's large direct and indirect effects as well as saturation's amplification of both effects is the municipal-level strategies of other actors with the capacity to influence voters *en masse*. Most plausibly, the large increases in incumbent vote share in Q1/2 in high saturation municipalities could result from incumbent parties or media outlets incorporating information from the ASF reports into large-scale campaign activities or news reports in response to the Facebook ad campaigns. The scope to respond to the intervention in this way was limited because the ads were distributed in the last week of the electoral campaign; the results of the ASF reports were also already available to parties and media outlets before Borde Político's intervention. Nevertheless, we assess these alternative explanations by examining online politician engagement with Borde Político's campaign and reporting on the ASF reports (and corruption more generally) by local newspapers. In each case, we detect little evidence of either alternative mechanism.

6.2.1 Online campaign responses to the Facebook ads

Online campaigning has become relatively common in Mexico; ads on the internet constituted an average of 7% of municipal campaign expenditures during the 2018 election campaigns. Candidates use social media platforms to announce their campaign promises, publicize their slogans, and denounce other candidates.

However, across all publicly accessible Facebook and Twitter accounts that we identified as belonging to candidates in our sample of municipalities, we found only four responses to Borde Político's information campaign on Facebook and none on Twitter. We additionally scraped thousands of comments, reactions, and shares relating to all Borde Político's Facebook ads and pages. Again, we observed negligible activity among Facebook users identified as running for other offices or appearing to be operatives in a municipal campaign. Section A.15 explains our data collection and summarizes the four reactions we found. Finally, none of the ads referencing the ASF in Facebook's library of election campaign ads (from May 7, 2018 until election day) reference the audit reports disseminated by Borde Político. It is highly unlikely that these scattered responses could account for the changes in voting behavior at the scale we observe.

6.2.2 Local media reporting of the ASF audits after the ad campaign

To examine whether media reporting related to Borde Político's Facebook ad campaign could explain the campaign's effects, we collected online data from 263 local newspapers that serve 92 of the 124 municipalities in our final sample (see Section A.16 for more details). The majority of local newspapers in Mexico provide significant amounts of content—including from their print editions—on their websites, often including full versions of the print editions, and local newspapers are still considered an important source of news content for local broadcast media outlets (Larreguy, Lucas and Marshall 2020). Over the 10 days between the start of the Facebook ad campaigns and the election, we searched for a variety of terms related to the specific content of the Facebook ads as well as more general references to corruption that the ad campaign might have inspired by increasing the demand for or stimulating the supply of such content. We estimate the effect of the campaign on such media reports using municipal-level regressions analogous to equation (4).[40]

We find little evidence to suggest that Borde Político's Facebook ad campaign induced a response from local media outlets. First, we were unable to detect any newspaper articles referencing the Facebook campaign, Borde Político's dissemination of information, or the ASF's reports between ad dissemination and election day. Given that media outlets do sometimes report on the outcomes of ASF audits (Larreguy, Marshall and Snyder 2020), this suggests that our sample of newspapers were unaware of the Facebook ads, had already reported on the ASF reports, or lacked incentives or resources to report on the issue during elections campaigns that covered all elected offices. Second, the ad campaigns also did not significantly increase reporting on issues related to corruption more generally. In particular, Appendix Table A22 reports that neither total mentions of corruption nor circulation-weighted mentions of corruption by local newspapers were systematically affected by the presence of a municipal Facebook ad campaign, the level of irregularities reported, or the campaign's saturation.

Finally, we consider whether the Facebook ads complement local news content more generally. To do so, we use data from Larreguy, Marshall and Snyder (2020) to examine heterogeneity in treatment effects by the number of local and non-local radio and television outlets covering an electoral precinct. Local outlets are distinguished by being located in the same municipality as a given precinct and producing local content. The results, reported in Appendix Tables A23 and A24, provide no evidence to suggest that either local or non-local media outlets amplified the effects of Facebook ads. If anything, media outlets may substitute for Facebook ads, potentially because some outlets had already covered ASF audit results or because their other content decreased the salience of the Facebook ads. Regardless, these findings reinforce the conclusion that the changes in vote choice induced by the Facebook ads were not driven by media coverage of the campaign.

[40] Since our randomization ensures that newspaper circulation is orthogonal to municipal treatment assignment and some small municipalities may not have local newspapers, we retain all municipalities for this analysis.

7 Conclusions

We show that non-partisan information campaigns on social media can support electoral accountability, particularly when the campaign activates social amplification mechanisms by targeting a large fraction of the electorate. Specifically, we demonstrate that Borde Político's large-scale pre-election Facebook ad campaign in Mexico substantially increased the vote share of municipal incumbent parties reported to have engaged in zero or negligible irregularities. The substantial direct and indirect effects of the campaign were greatest where the campaign targeted 80% rather than 20% of the electorate, and appear to be driven primarily by tacit or explicit interactions between voters across the municipality that were induced by the campaign, rather than responses by—or complementarities with—political campaigns or other media outlets.

Showing that the electoral impacts of factual non-partisan messages on Facebook spread through social networks implies that low-cost interventions on social media platforms can support the foundations of democratic accountability. This is particularly important in the Global South, where social media use is growing rapidly and significant governance challenges remain. Neverthelessfurther research is required to understand whether the effects generalize to contexts where government malfeasance was less salient and how to maximize the impact of factual non-partisan campaigns in competitive online markets for political content. Evidence of similarly large effects of a Facebook ad campaign encouraging voters to report electoral irregularities in Colombia (Garbiras-Diaz and Montenegro forthcoming) cements our reason for optimism. Both studies suggest that Facebook ad campaigns costing just several hundred US dollars in the average municipality are underutilized, although partisan campaigns in the Global North have not produced such dramatic results (Hager 2019; Liberini et al. 2020). Since online information campaigns can have substantial electoral impacts and these tools are equally available to more malfeasant actors, our results also help to inform debate surrounding—and the possible trade-offs underpinning—whether regulation is required to ensure that elections are not hijacked by misinformation.

Our findings also suggest that information campaign saturation—the share of a market targeted by an information campaign—can help explain the greater impact of information dissemination conducted by mass-reach broadcast and social media outlets. Appendix Table A1 more systematically documents this correlation across 15 studies that estimated the effect of incumbent performance information on voting behavior. Of course, media outlets might do more than just provide access to information for large audiences; they can distort, filter, and frame content in different ways (see Prat and Strömberg 2013; Strömberg 2015). Future research should more directly establish the extent to which saturation and these other potential features drive the role of mass media in promoting or hindering electoral accountability. Finer-grained distinctions between the mechanisms that shape voter responses—possibly including beliefs updating, priming, or tacit and explicit coordination—may represent an important topic for further study, especially for optimizing the impact of non-partisan social media campaigns.

References

Adida, Claire, Jessica Gottlieb, Eric Kramon and Gwyneth McClendon. 2020. “When does information influence voters? the joint importance of salience and coordination.” *Comparative Political Studies* 53(6):851–891.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer and Matthew Gentzkow. 2020. “The Welfare Effects of Social Media.” *American Economic Review* 110(3):629–676.

Alt, James E., Amalie Jensen, Horacio Larreguy, David D. Lassen and John Marshall. 2022. “Diffusing political concerns: How unemployment information passed between social ties influences Danish voters.” *Journal of Politics* 84(1):383–404.

Arias, Eric, Horacio Larreguy, John Marshall and Pablo Querubín. 2018. “Does the Content and Mode of Delivery of Information Matter for Electoral Accountability? Evidence from a Field Experiment in Mexico.” Working paper.

Arias, Eric, Horacio Larreguy, John Marshall and Pablo Querubín. forthcoming. “Priors rule: When do Malfeasance Revelations Help or Hurt Incumbent Parties?” *Journal of the European Economic Association* .

Arias, Eric, Pablo Balán, Horacio Larreguy, John Marshall and Pablo Querubín. 2019. “Information Provision, Voter Coordination, and Electoral Accountability: Evidence from Mexican Social Networks.” *American Political Science Review* 113(2):475–498.

Auditoría Superior de la Federación. 2014. “Informe del Resultado de la Fiscalización Superior de la Cuenta Pública 2012.” Audit Summary Report.

Avis, Eric, Claudio Ferraz and Frederico Finan. 2018. “Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians.” *Journal of Political Economy* 126(5):1912–1964.

Baird, Sarah, J. Aislinn Bohren, Craig McIntosh and Berk Özler. 2018. “Optimal Design of Experiments in the Presence of Interference.” *Review of Economics and Statistics* 100(5):844–860.

Banerjee, Abhijit, Nils Enevoldsen, Rohini Pande and Michael Walton. 2020. “Public Information is an Incentive for Politicians: Experimental Evidence from Delhi Elections.” NBER working paper 26925.

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

Berry, Christopher R. and William G. Howell. 2007. "Accountability and Local Elections: Rethinking Retrospective Voting." *Journal of Politics* 69(3):844–858.

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

Besley, Timothy and Robin Burgess. 2002. "The Political Economy of Government Responsiveness: Theory and Evidence from India." *Quarterly Journal of Economics* 117(4):1415–1451.

Bhandari, Abhit, Horacio Larreguy and John Marshall. forthcoming. "Able and mostly willing: An empirical anatomy of information's effect on voter-driven accountability in Senegal." *American Journal of Political Science* .

Bidwell, Kelly, Katherine Casey and Rachel Glennerster. 2020. "Debates: Voting and expenditure responses to political communication." *Journal of Political Economy* 128(8):2880–2924.

Boas, Taylor C., F. Daniel Hidalgo and Marcus A. Melo. 2019. "Norms versus Action: Why Voters Fail to Sanction Malfeasance in Brazil." *American Journal of Political Science* 63(2):385–400.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D.I. Kramer, Cameron Marlow, Jaime E. Settle and James H. Fowler. 2012. "A 61-million-person experiment in social influence and political mobilization." *Nature* 489:295–298.

Bowles, Jeremy and Horacio Larreguy. 2020. "Who Debates, Wins? At-Scale Experimental Evidence on Debate Participation in a Liberian Election." Working paper.

Braghieri, Luca, Ro'ee Levy and Alexey Makarin. 2021. "Social Media and Mental Health." Working paper.

Breza, Emily, Fatima Cody Stanford, Marcella Alsan, Burak Alsan, Abhijit Banerjee, Arun G. Chandrasekhar, Sarah Eichmeyer, Traci Glushko, Paul Goldsmith-Pinkham, Kelly Holland, Emily Hoppe, Mohit Karnani, Sarah Liegl, Tristan Loisel, Lucy Ogbu-Nwobodo, Benjamin A. Olken, Carlos Torres, Pierre-Luc Vautrey, Erica T. Warner, Susan Wootton and Esther Duflo. 2021. "Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: a cluster randomized controlled trial." *Nature Medicine* 27(9):1622–1628.

Buntaine, Mark T., Ryan Jablonski, Daniel L. Nielson and Paula M. Pickering. 2018. "SMS texts on corruption help Ugandan voters hold elected councillors accountable at the polls." *Proceedings of the National Academy of Sciences* 115(26):6668–6673.

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov and Maria Petrova. 2019. "Social Media and Xenophobia: Evidence from Russia." NBER working paper 26567.

Caprettini, Bruno, Marcel Caesmann, Hans-Joachim Voth and David Yanagizawa-Drott. 2021. "Going Viral: Propaganda, Persuasion and Polarization in 1932 Hamburg." Working paper.

Cavalcanti, Francisco, Gianmarco Daniele and Sergio Galletta. 2018. "Popularity shocks and political selection." *Journal of Public Economics* 165:201–216.

Chong, Alberto, Ana De La O, Dean Karlan and Leonard Wantchekon. 2015. "Does Corruption Information Inspire the Fight or Quash the Hope? A Field Experiment in Mexico on Voter Turnout, Choice and Party Identification." *Journal of Politics* 77(1):55–71.

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

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

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot and Philippe Zamora. 2013. "Do labor market policies have displacement effects? Evidence from a clustered randomized experiment." *Quarterly Journal of Economics* 128(2):531–580.

Cruz, Cesi, Philip Keefer and Julien Labonne. 2021. "Buying informed voters: New effects of information on voters and candidates." *Economic Journal* 131(635):1105–1134.

Cruz, Cesi, Philip Keefer, Julien Labonne and Francesco Trebbi. 2019. "Making Policies Matter: Voter Responses to Campaign Promises." Working paper.

de Figueiredo, Miguel F.P., F. Daniel Hidalgo and Yuri Kasahara. forthcoming. "When Do Voters Punish Corrupt Politicians? Experimental Evidence from a Field and Survey Experiment." *British Journal of Political Science* pp. 1–12.

De La O, Ana, L. 2013. "Do Conditional Cash Transfers Affect Electoral Behavior? Evidence from a Randomized Experiment in Mexico." *American Journal of Political Science* 57(1):1–14.

DellaVigna, Stefano and Matthew Gentzkow. 2010. "Persuasion: Empirical Evidence." *Annual Review of Economics* 2(1):643–669.

Druckman, James N., Matthew S. Levendusky and Audrey McLain. 2018. "No need to watch: How the effects of partisan media can spread via interpersonal discussions." *American Journal of Political Science* 62(1):99–112.

Duflo, Esther and Emmanuel Saez. 2003. "The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment." *Quarterly Journal of Economics* 118(3):815–842.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan Hyde, Craig McIntosh and Gareth Nellis. 2019. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*. Cambridge University Press.

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

Fearon, James D. 1999. Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance. In *Democracy, Accountability, and Representation*, ed. Adam Przeworski, Susan Stokes and Bernard Manin. Cambridge University Press.

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50(1):5–25.

Fergusson, Leopoldo. 2014. "Media markets, special interests, and voters." *Journal of Public Economics* 109(C):13–26.

Ferraz, Claudio and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123(2):703–745.

Finan, Frederico, Enrique Seira and Alberto Simpser. 2021. "Voting with one's neighbors: Evidence from migration within Mexico." *Journal of Public Economics* 202:104495.

Fujiwara, Thomas, Karsten Müller and Carlo Schwarz. 2020. "The Effect of Social Media on Elections: Evidence from the United States." Working paper.

Garbiras-Diaz, Natalia and Mateo Montenegro. forthcoming. "All Eyes on Them: A Field Experiment on Citizen Oversight and Electoral Integrity." *American Economic Review* .

George, Siddharth, Sarika Gupta and Yusuf Neggers. 2019. "Can We Text Criminal Politicians Out of Office? Evidence from a Mobile Experiment in India." Working paper.

Gilbert, Eric and Karrie Karahalios. 2009. Predicting tie strength with social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. pp. 211–220.

Grossman, Guy and Kristin Michelitch. 2018. "Information Dissemination, Competitive Pressure, and Politician Performance between Elections: A Field Experiment in Uganda." *American Political Science Review* 112(2):280–301.

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

Hager, Anselm. 2019. "Do online ads influence vote choice?" *Political Communication* 36(3):376–393.

Hootsuite and We are Social. 2018*a*. "2018 Digital Yearbook.". [Online; accessed 9-June-2018].
URL: *https://hootsuite.com/pages/digital-in-2018*

Hootsuite and We are Social. 2018*b*. "Digital in 2018 in Mexico.". [Online; accessed 9-June-2018].
URL: *https://hootsuite.com/pages/digital-in-2018*

Horvitz, Daniel G. and Donovan J. Thompson. 1952. "A generalization of sampling without replacement from a finite universe." *Journal of the American Statistical Association* 47(260):663–685.

Humphreys, Macartan and Jeremy Weinstein. 2012. "Policing Politicians: Citizen Empowerment and Political Accountability in Uganda Preliminary Analysis." Working paper.

Iyengar, Shanto and Donald R. Kinder. 1987. *News That Matters: Television and American Opinion*. University of Chicago Press.

Jones, Jason J., Robert M. Bond, Eytan Bakshy, Dean Eckles and James H. Fowler. 2017. "Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 US presidential election." *PloS One* 12(4):e0173851.

Katz, Elihu and Paul Felix Lazarsfeld. 1955. *Personal Influence: The part played by people in the flow of mass communications*. Free Press.

Kendall, Chad, Tommaso Nannicini and Francesco Trebbi. 2015. "How Do Voters Respond to Information? Evidence from a Randomized Campaign." *American Economic Review* 105(1):322–53.

Khemani, Stuti, Ernesto Dal Bó, Claudio Ferraz, Frederico S. Finan, Johnson Stephenson, Louise Corinne, Adesinaola M. Odugbemi, Dikshya Thapa and Scott D. Abrahams. 2016. "Making Politics Work for Development: Harnessing Transparency and Citizen Engagement." World Bank Policy Research Report 106337.

Laakso, Markku and Rein Taagepera. 1979. ""Effective" number of parties: a measure with application to West Europe." *Comparative Political Studies* 12(1):3–27.

Langston, Joy. 2003. "Rising from the Ashes?: Reorganizing and Unifying the PRI's State Party Organizations after Electoral Defeat." *Comparative Political Studies* 36(3):293–318.

Larreguy, Horacio, Christopher Lucas and John Marshall. 2020. “When do Media Stations Support Political Accountability? A Field Experiment in Mexico.” Working paper.

Larreguy, Horacio and Joana C. M. Monteiro. 2019. “Media Networks and Political Accountability: Evidence from Radio Networks in Brazil.” Working paper.

Larreguy, Horacio, John Marshall and Jr. Snyder, James M. 2020. “Publicising Malfeasance: When the Local Media Structure Facilitates Electoral Accountability in Mexico.” *Economic Journal* 130(631):2291–2327.

Lazarsfeld, Paul F., Bernard Berelson and Hazel Gaudet. 1944. *The People’s Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. Columbia University Press.

Levy, Ro’ee. 2021. “Social Media, News Consumption, and Polarization: Evidence from a Field Experiment.” *American Economic Review* 111(3):831–870.

Liberini, Federica, Michela Redoano, Antonio Russo, Angel Cuevas and Ruben Cuevas. 2020. “Politics in the Facebook Era: Evidence from the 2016 US Presidential Elections.” Working paper.

Lierl, Malte and Marcus Holmlund. 2019. “Can Scorecards Improve Municipal Government Performance? A Field Experiment.” Working paper.

Marshall, John. 2022. “Tuning in, voting out: News consumption cycles, homicides, and electoral accountability in Mexico.” Working paper.

McKenzie, David. 2012. “Beyond baseline and follow-up: The case for more T in experiments.” *Journal of Development Economics* 99(2):210–221.

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

Müller, Karsten and Carlo Schwarz. 2021. “Fanning the flames of hate: Social media and hate crime.” *Journal of the European Economic Association* 19(4):2131–2167.

Pande, Rohini. 2011. “Can Informed Voters Enforce Better Governance? Experiments in Low-Income Democracies.” *Annual Review of Economics* 3(1):215–237.

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.

Poushter, Jacob, Caldwell Bishop and Hanyu Chwe. 2018. “Social Media Use Continues to Rise in Developing Countries but Plateaus Across Developed Ones.” *Pew Research Center* 22.

Prat, Andrea and David Strömberg. 2013. The Political Economy of Mass Media. In *Advances in Economics and Econometrics: Volume 2, Applied Economics: Tenth World Congress*. Vol. 50 Cambridge University Press p. 135.

Reinikka, Ritva and Jakob Svensson. 2011. "The power of information in public services: Evidence from education in Uganda." *Journal of Public Economics* 95(7-8):956–966.

Rodríguez-Castelán, Carlos, Kiyomi Cadena and Laura Moreno. 2018. "Evaluación de largo plazo de los impactos distributivos del Fondo de Infraestructura Social Municipal (FISM)." Working paper.

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

Sinclair, Betsy, Margaret McConnell and Donald P. Green. 2012. "Detecting spillover effects: Design and analysis of multilevel experiments." *American Journal of Political Science* 56(4):1055–1069.

Snyder, James M., Jr. and David Strömberg. 2010. "Press Coverage and Political Accountability." *Journal of Political Economy* 118(2):355–408.

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

Strömberg, David. 2015. "Media and Politics." *Annual Review of Economics* 7:173–205.

Yanagizawa-Drott, David. 2014. "Propaganda and conflict: Evidence from the Rwandan genocide." *Quarterly Journal of Economics* 129(4):1947–1994.

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

A Appendix

Contents

A.1 Mixed evidence that information has improved electoral accountability

A substantial body of credible evidence evaluating the impact of information on electoral accountability in the Global South—where political accountability often remains limited (e.g. Khemani et al. 2016; Pande 2011)—has accumulated over more than a decade.[1] Table A1 summarizes the results of studies that leverage either field or natural experiments to estimate the effects of providing information documenting at least one aspect of incumbent performance to voters before elections on administrative or self-reported measures of votes for the incumbent or the incumbent party.[2] We distinguish studies in terms of: (i) whether the findings are broadly consistent with standard theories of electoral accountability, by which we mean providing evidence that signals of good (bad) performance or signals that exceed (fall below) prior expectations increase (decrease) support for the incumbent candidate or party;[3] and (ii) whether treatment saturation—the fraction of the electorate that has direct access to information about a given incumbent politician or party—was low (targeting less than 10% of the electorate represented by the incumbent about which information was provided), medium (10-40% of the electorate), or high (greater than 40% of the electorate). Appendix A.1.1 discusses the classification of particular studies in greater detail.

As Table A1 shows, the information campaigns most likely to support electoral accountability involve higher levels of saturation. Most of the studies examining high saturation campaigns leverage spatial variation in access to media outlets that are likely to report on local incumbents' performance. Ferraz and Finan's (2008) finding that the sanctioning of incumbents based on the outcomes of independent audit reports in Brazil is driven by municipalities with access to a local radio station, while Larreguy, Marshall and Snyder's (2020) results—further exploiting plausibly exogenous variation in access to local media—suggest that voter rewards for and punishment of malfeasance revelations from a similar audit program in Mexico are caused by the media. Marshall (2022) also finds that access to local media in Mexico drives sanctioning of municipal incumbents overseeing spikes in homicide rates before elections. While we focus on voter responses to information provided before elections in this article, other studies show that politician behavior in office responds to the presence of the media (e.g. Avis, Ferraz and Finan 2018; Besley and Burgess 2002;

[1]For studies covering the Global North, see e.g. Berry and Howell (2007), Fergusson (2014), Kendall, Nannicini and Trebbi (2015), and Snyder and Strömberg (2010).

[2]We exclude studies exclusively measuring vote intentions, on the basis that they are normally elicited shortly after surveys, may be vulnerable to social desirability biases, and are hypothetical by construction. We also exclude recent articles examining debates between candidates (e.g. Bowles and Larreguy 2020; Bidwell, Casey and Glennerster 2020; Platas and Raffler 2021), on the basis that a wide range of information—which may or may not include incumbent performance indicators—is provided.

[3]This categorization is not always straightforward. For example, defining good and bad levels of performance in Ferraz and Finan (2008) and Larreguy, Marshall and Snyder (2020) is not obvious because prior beliefs were not registered, although we believe in such cases the heterogeneity with respect to performance paints a clear picture of support for electoral accountability. Studies that did not provide information about incumbent performance (e.g. Cruz, Keefer and Labonne 2021; Cruz et al. 2019) were excluded.

Larreguy and Monteiro 2019; Reinikka and Svensson 2011) or the threat of future media reporting (Banerjee et al. 2020; Grossman and Michelitch 2018).

Although they do not vary information saturation levels, a number of medium and high saturation field experiments also provide compelling evidence consistent with electoral accountability. Randomizing the mass provision of incumbent performance information shortly before elections, these studies show—across a variety of settings—that voters reward incumbents at the ballot box for exerting greater legislative effort (Banerjee et al. 2011) and engaging in less malfeasance than expected (Arias et al. forthcoming), and also sanction politicians accused of severe crimes like murder (George, Gupta and Neggers 2019). Several medium saturation studies also find some evidence that voters support (reject) better(worse)-performing incumbents, but observe different degrees of electoral accountability across different layers of government (Buntaine et al. 2018) or find that voters primarily sanction challengers (Chong et al. 2015; de Figueiredo, Hidalgo and Kasahara forthcoming).

In contrast, many low saturation interventions fail to detect effects of incumbent performance information on voting behavior. Indeed, randomized information provision pertaining to policy decisions and legislative effort did little to influence voters in northern Brazil (Boas, Hidalgo and Melo 2019), Burkina Faso (Lierl and Holmlund 2019), or Uganda (Humphreys and Weinstein 2012). The first two studies come from the recent Metaketa initiative, which coordinated similar accountability experiments across six countries in the Global South and found negligible effects on average (Dunning et al. 2019).[4] Adida et al.'s (2020) study is an exception with mixed findings showing that legislator performance that exceeded expectations was punished in a very low saturation version of the intervention and, when combined with civics training, was rewarded in a somewhat higher saturation version of the same intervention. The single low saturation field experiment to report that incumbent performance consistently impacted vote choices comes from rural Senegal (Bhandari, Larreguy and Marshall forthcoming), where there is evidence of substantial information diffusion within treated villages.

In sum, our review of extant studies indicates that saturation may moderate the impact of information dissemination on electoral accountability. However, the notable correlation in Table A1 is, at best, suggestive of a causal relationship. First, saturation is likely correlated with many potential confounds, including the type of content provided, how prominently and persuasively information is communicated, or the predisposition of voters in a given context to respond. Second, studies examining high saturation information provision have generally leveraged observational data from natural experiments that may only be published when significant results are found. Since such designs may be under-powered to detect small effects or more vulnerable than pre-registered ex-

[4] The Metaketa studies providing incumbent performance information are: Adida et al. (2020), Arias et al. (forthcoming), Boas, Hidalgo and Melo (2019), Buntaine et al. (2018), and Lierl and Holmlund (2019); one intervention in India was withdrawn, while another focused on debates rather than performance.

Table A1: The electoral effects of providing incumbent performance information, by treatment saturation

		Saturation of information dissemination		
		Low (0-10%)	*Medium (10-40%)*	*High (40-100%)*
Consistency of findings with theories of electoral accountability	*Mostly null findings*	Boas, Hidalgo and Melo (2019), Humphreys and Weinstein (2012), Lierl and Holmlund (2019).		
	Mixed findings	Adida et al. (2020).	Buntaine et al. (2018), Chong et al. (2015), de Figueiredo, Hidalgo and Kasahara (forthcoming).	
	Mostly consistent findings	Bhandari, Larreguy and Marshall (forthcoming).	Arias et al. (forthcoming), Banerjee et al. (2011), George, Gupta and Neggers (2019).	Ferraz and Finan (2008), Larreguy, Marshall and Snyder (2020), Marshall (2022).

Notes: Underlined articles leverage experimental variation in access to information. See Section A.1.1 for more details on the classification of articles and findings.

periments to remain in the "file drawer," the observed cross-study correlation could instead reflect publication biases. Third, the one study—by Adida et al. (2020)—that has experimentally varied electorate-level saturation introduced only limited variation in saturation and no experimental condition involved high levels of saturation.[5] Moreover, only 30 commune electorates were randomized into higher or lower saturation in this study. Our intervention also differs in terms of delivering information via mass media, rather than communal projector viewings or in private via video at the citizen's home.

A.1.1 Classification of extant informational interventions

As noted above, we reviewed field and natural experimental studies estimating the effect of providing voters with incumbent performance information in the Global South. We thus excluded studies in the Global North and studies in the Global South that provided non-performance information (e.g. candidate debates). We also excluded studies that only contained self-reported outcomes measured immediately after treatment. As noted above, saturation is defined by the share of voters with *direct access* to incumbent performance information. We, therefore, do not count untreated or indirectly treated voters within treated clusters in our computations of saturation.

Below we summarize the studies included in Table A1 in alphabetical order (with the natural experiments grouped at the end), and discuss our coding decisions:

1. In Adida et al. (2020), we code the borderline statistically significant negative effect of positive information on incumbent vote share in villages within low-dosage communes in column (2) of Table 4 as "mostly null findings" because the direction goes against the expectations of standard electoral accountability models. The borderline statistically significantly positive effect of providing better-than-expected incumbent performance information in villages within high-dosage communes is coded as "mixed findings," given that this effect only holds when incumbent performance information is accompanied by civic training. A typical commune contains around 50 villages. We code their low-dosage case, where 2 villages per commune (or around 4%) were treated, as "low" saturation. We also code their high-dosage case, where 15 villages per commune (or 30%) were treated, as "low" saturation because only 12-15% of households within each treated village—and thus 4.5% of voters within high-dosage communes—had direct access to treatment. Overall, we code this study as providing "mixed findings."

2. In Arias et al. (forthcoming), we code the significant positive effect of the information treatment for 0% malfeasance and the significantly negative interaction between treatment and

[5] Adida et al. (2020) compared 15 low-dosage communes with 15 higher-dosage communes, where 12-15% of voters were directly treated within around 2% and 27% of villages within each commune respectively. The high-dosage treatment thus reached around 4.5% of citizens within high-dosage communes.

incumbent malfeasant spending on incumbent party vote share, both from column (4) of Table 4, as "mostly consistent findings." The campaign's saturation is classified as "medium" because around 20% of precincts were treated in a typical municipality, although only up to 200 leaflets were distributed to households (more than half of the households) within a given electoral precinct.

3. In Banerjee et al. (2011), we code the significant positive effects of the interaction between treatment and overall incumbent quality in column (4) and incumbent performance in column (5) of Table 4 on incumbent party vote share "mostly consistent findings." We code saturation as "medium" due to the fact that 200 polling stations—each with roughly 1,000 voters—are treated out of 775 selected polling stations in ten constituencies with high slum density—each with approximately 100,000 citizens.

4. In Bhandari, Larreguy and Marshall (forthcoming), we code the significant positive interaction effect between the information treatment and the incumbent local performance index on validated reported vote for the incumbent in Table 8 and polling station-level incumbent party vote share in Table 9 as "mostly consistent findings." Saturation is coded as "low:" although 375 villages were treated across 5 constituencies in Senegal, each containing approximately 300 villages, only 9 voters per village were directly treated.

5. In Boas, Hidalgo and Melo (2019), we code the zero treatment effects of reporting either approved or rejected account through a field experiment reported in Figure 3 as "mostly null findings." Saturation is coded as "low" because the information was randomly distributed to 1,600 registered voters across 47 municipalities in Brazil.

6. In Buntaine et al. (2018), we code the zero effects of either good or bad news about incumbent performance on chairman vote and the statistically significantly positive (negative) effects of good (bad) news about incumbent performance on councilor vote as "mixed findings." Saturation is coded as "medium" because messages were sent to 16,083 citizens in 762 villages and we expect these to cover a medium share of the villages under the councilors in the experimental sample. This study also varied saturation across villages, as opposed to across electorates, but found little evidence to suggest an impact of localized saturation (see Figure 2).

7. In Chong et al. (2015), we code the effects of the treatments as "mixed." Table 4 shows that the provision of incumbent corruption information reduced turnout for the incumbent and challenger where high levels of incumbent corruption (top tercile) were reported. However, such effects are greater for the challenger than the incumbent, which suggest that overall corruption revelations favored corrupt incumbents. Saturation is coded as "medium" because,

although all electoral precincts in each of the 12 sample municipalities were assigned to one of four treatment conditions (including a pure control), the authors ended up pooling three of these conditions as a control group, leaving a quarter of precincts per municipality treated by their definition.

8. In de Figueiredo, Hidalgo and Kasahara (forthcoming), we code the findings as "mixed" based on the significant negative effect of information showing the challenger to be corrupt on vote share for the challenger in Table 2 and the insignificant positive effect of information also showing the incumbent to be corrupt on vote share for the incumbent in Table 3. Saturation is coded as "medium" on the basis that 187,177 fliers with candidate information were delivered to 200 out of 1,759 precincts in Sao Paulo.

9. In George, Gupta and Neggers (2019), we code the significant negative effects of revelations of candidate murder-related charges on candidate vote share in Table 6 as "mostly consistent findings." We code saturation as "medium" because the authors treated 80% of an experimental sample of 4,131 villages, out of a total of 9,627 villages from 39 assembly constituencies.

10. In Humphreys and Weinstein (2012), we code the null treatment effect on reported vote of reporting relatively good news about incumbent performance relative to the prior beliefs in Table 3 as "mostly null findings." Saturation is coded as "low" since, despite the many information dissemination efforts, treatment information reached a very small share of the electorate. Specifically, the authors undertook two main information dissemination efforts prior to the election: community-wide workshops and blanket treatment of polling stations with scorecard results. However, while workshop attendance averaged about 120 people and 1,500 copies of the local language scorecard were handed out to be shared more broadly in each workshop, only one workshop was conducted per constituency, in a context where constituencies average around 50,000 voters each. Similarly, only voters in two polling station areas per constituency were assigned to receive scorecard results.

11. In Lierl and Holmlund (2019), we code the null treatment effects of both good and bad information about incumbent performance reported in Table 3 as "mostly null effects." Saturation is coded as "low" because the information was only randomly distributed to 752 study participants across 38 municipalities in Burkina Faso.

12. Finally, Ferraz and Finan (2008), Larreguy, Marshall and Snyder (2020), and Marshall (2022) are all articles that analyze the effect of media revelation of incumbent malfeasance or municipal violence on incumbent vote share, finding large sanctioning or reward effects that reflect performance (of some form) in each case; we thus code each as showing "mostly consistent

findings." Due to the large coverage or circulation of the media outlets that they focus on, the saturation in each study is coded as "high."

A.2 Infographics on Borde Político's Facebook ad pages

Figures A1a or A1b show examples of the infographic available on the Facebook page associated with the ads, with the former reporting 0% irregularities and the latter type reporting greater than 0% irregularities.

A.3 Municipalities included in the sample

Figure A2 reports the 128 municipalities included in the sample. The criteria for inclusion are described in the main paper. The two delegaciones in Mexico City that were audits were excluded from our sample because such delegaciones operated differently from municipalities during the relevant time period.

A.4 Engagement with Facebook ads

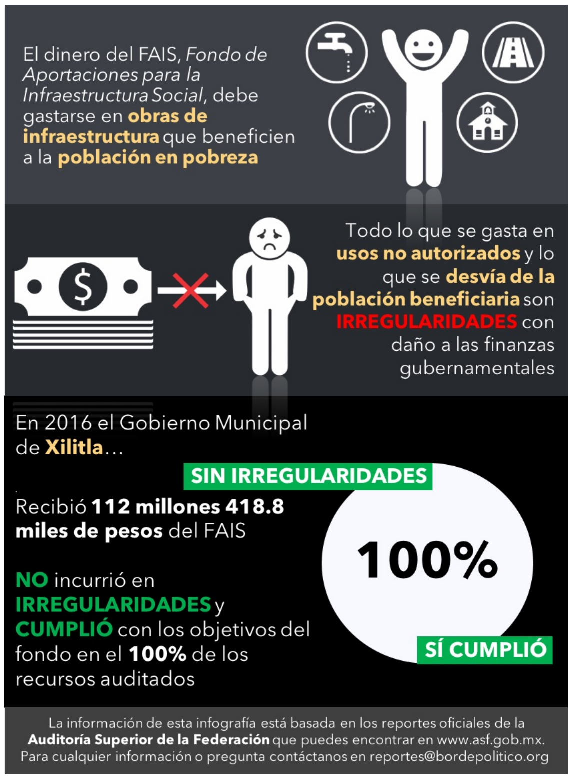
Figure A3 plots trends in Facebook user engagement with Borde Político's Facebook ad campaign over its duration. Figure A4 further reports the distribution of aggregate engagement with the Facebook ads within low and high saturation municipalities. We were not able to obtain day-by-day data on watching at least 10 seconds of the ad.

We also use Facebook's analytics data to understand the demographic characteristics of the Facebook users that engaged with Borde Político's ads. For this analysis, we focus on two measures of engagement that are available for all demographic subgroups: the share of unique viewers from a particular demographic and the share of views of at least 3 seconds within a given subgroup.

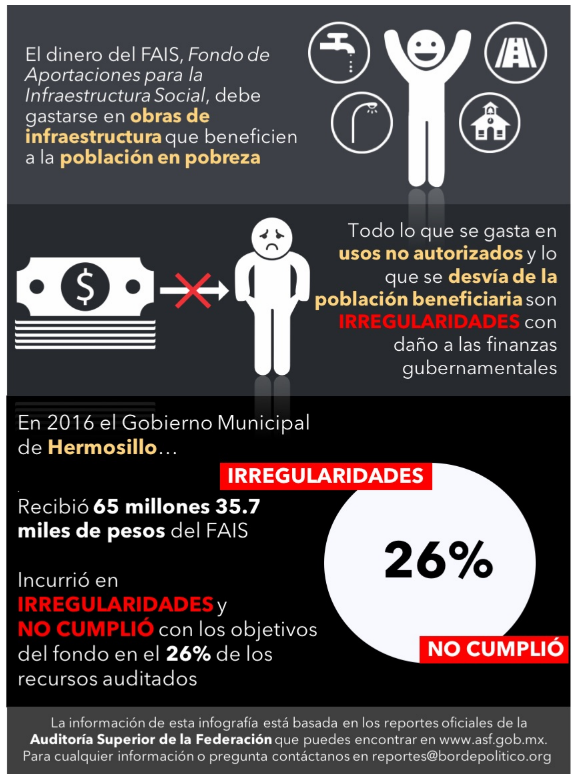
The outcome means at the foot of Table A2 characterize the types of Facebook users that received and watched the ads in low saturation municipalities. First, column (1) of panels A and C shows that the ads reached and were watched by men and women in roughly equal proportion. Second, columns (2)-(7) of panel A indicate that the ads disproportionately reached younger adults. However, panel C shows that the share of Facebook users that watched at least 3 seconds of the ad was broadly in line with the 2010 Census adult age distribution. This suggests that younger Facebook users were relatively less likely to watch the ad when it appeared, but more likely to receive the ad. Third, columns (8)-(11) show that ad consumption increases through night (12pm-6am), morning (6am-12pm), afternoon (12pm-6pm), and evening (6pm-12am), with the evening period registering greatest ad reach and views of at least 3 seconds. Finally, columns (12) and (13) demonstrate that around 90% of engagement occurred via a smartphone.

Figure A1: Examples of treatment infographics

(a) Example of an infographic from a municipality (Xilitla, San Luis Potosí) where irregularities were 0%



(b) Example of an infographic from a municipality (Hermosillo, Sonora) where irregularities were above 0%



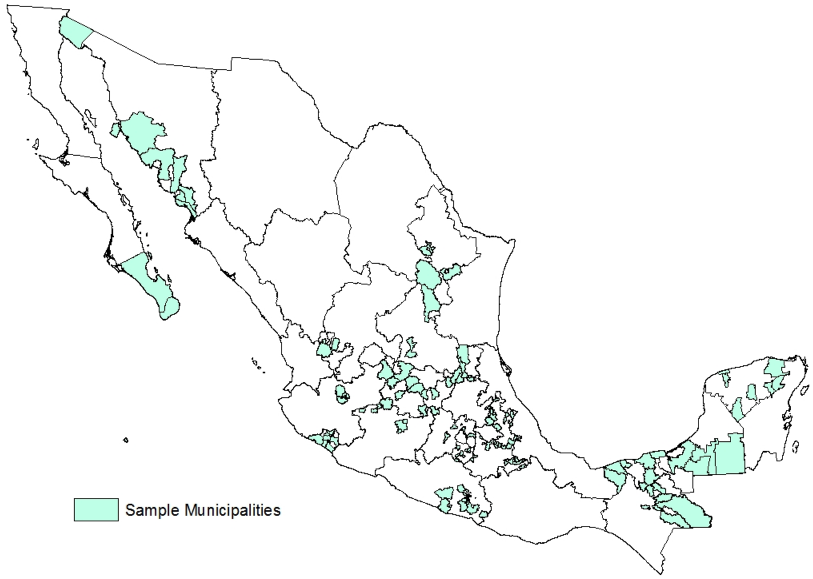
Notes: In English, the first panel of the left infographic says "The funds from the FAIS, the Fund for Social Infrastructure, must be spent on infrastructure projects benefiting the poor;" the second panel says "All funds spent on unauthorized projects or not benefiting the poor constitute irregularities harming public finances;" the third panel says "In 2016, the municipal government of Xilitla received 112 million 419.8 thousand pesos" and "It did not incur in any regularity fulfilling the intended purposes of the fund in 100% of the audited funds;" and the fourth panel says "The information from the infographic is from the ASF's official audit reports that can be accessed at asf.gob.mx" and "To request more information or make an inquiry, you can contact us at reportes@bordepolitico.org." The first, second, and last panel of the right infographic do not change, and the third panel says "In 2016, the municipal government of Hermosillo received 65 million 35.7 thousand pesos" and "It incurred in regularities not fulfilling the intended purposes of the fund in 26% of the audited funds."

Table A2: Engagement with Facebook ads in treated municipalities, by demographic subgroup

	Consumption of Facebook ads by demographic group												
	% female	% aged 18-24	% aged 25-34	% aged 35-44	% aged 45-54	% aged 55-64	% aged 65+	% morning	% afternoon	% evening	% night	% smartphone	% desktop
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Panel A: unique viewers													
High saturation	-0.016	-0.011	-0.009	0.001	0.005	0.008	0.005**	0.008	-0.010	0.004	-0.001	-0.015**	0.015***
	(0.015)	(0.015)	(0.007)	(0.006)	(0.008)	(0.005)	(0.002)	(0.008)	(0.008)	(0.005)	(0.002)	(0.007)	(0.005)
Observations	83	83	83	83	83	83	83	83	83	83	83	83	83
R^2	0.63	0.63	0.60	0.64	0.56	0.63	0.72	0.61	0.58	0.63	0.61	0.79	0.82
Low saturation outcome mean	0.49	0.33	0.32	0.19	0.10	0.04	0.02	0.27	0.31	0.37	0.05	0.93	0.07
Low saturation outcome std. dev.	0.09	0.07	0.04	0.03	0.04	0.03	0.01	0.04	0.04	0.03	0.01	0.04	0.04
Panel B: unique viewers, by irregularities quartile													
Q3	0.035	-0.009	-0.017	0.005	0.009	0.008	0.005	-0.019	0.017	0.006	-0.004	-0.009	0.015
	(0.063)	(0.036)	(0.015)	(0.014)	(0.017)	(0.013)	(0.006)	(0.020)	(0.020)	(0.010)	(0.007)	(0.019)	(0.015)
Q4	-0.007	-0.001	0.002	-0.018	-0.000	0.010	0.007	-0.000	0.018	-0.010	-0.007	-0.010	0.014
	(0.021)	(0.039)	(0.022)	(0.015)	(0.022)	(0.014)	(0.006)	(0.023)	(0.024)	(0.016)	(0.005)	(0.014)	(0.011)
High saturation	-0.019	-0.007	-0.012	-0.003	0.003	0.011	0.008*	0.014	-0.009	-0.004	-0.000	-0.026**	0.024**
	(0.016)	(0.027)	(0.010)	(0.010)	(0.014)	(0.010)	(0.004)	(0.013)	(0.013)	(0.006)	(0.002)	(0.012)	(0.009)
High saturation × Q3	-0.034	0.020	0.019	-0.002	-0.014	-0.015	-0.007	0.000	-0.006	0.003	0.003	0.020	-0.016
	(0.109)	(0.045)	(0.021)	(0.017)	(0.025)	(0.017)	(0.008)	(0.025)	(0.027)	(0.010)	(0.008)	(0.028)	(0.018)
High saturation × Q4	0.047**	-0.039	-0.001	0.018	0.022	0.002	-0.001	-0.028	0.005	0.029	-0.007	0.029	-0.020
	(0.018)	(0.047)	(0.024)	(0.017)	(0.026)	(0.018)	(0.007)	(0.026)	(0.019)	(0.023)	(0.006)	(0.018)	(0.013)
Observations	83	83	83	83	83	83	83	83	83	83	83	83	83
R^2	0.66	0.65	0.62	0.66	0.58	0.65	0.73	0.64	0.60	0.66	0.66	0.81	0.83
Low saturation outcome mean	0.49	0.33	0.32	0.19	0.10	0.04	0.02	0.27	0.31	0.37	0.05	0.93	0.07
Low saturation outcome std. dev.	0.09	0.07	0.04	0.03	0.04	0.03	0.01	0.04	0.04	0.03	0.01	0.04	0.04
Panel C: views (of 3 seconds)													
High saturation	-0.013	-0.009	-0.012	0.001	-0.000	0.012	0.010	0.008	-0.018**	0.010	0.000	-0.027**	0.019***
	(0.016)	(0.016)	(0.013)	(0.007)	(0.010)	(0.011)	(0.006)	(0.008)	(0.008)	(0.007)	(0.003)	(0.010)	(0.006)
Observations	83	83	83	83	83	83	83	83	83	83	83	83	83
R^2	0.63	0.59	0.60	0.55	0.55	0.64	0.66	0.62	0.62	0.66	0.55	0.82	0.80
Low saturation outcome mean	0.49	0.21	0.27	0.23	0.16	0.09	0.04	0.27	0.29	0.38	0.06	0.88	0.07
Low saturation outcome std. dev.	0.10	0.08	0.07	0.03	0.05	0.05	0.03	0.04	0.04	0.04	0.01	0.08	0.05
Panel D: views (of 3 seconds), by irregularities quartile													
Q3	0.040	0.025	-0.025	-0.025	0.005	0.009	0.010	-0.021	0.030	-0.001	-0.007	-0.002	0.006
	(0.064)	(0.055)	(0.035)	(0.015)	(0.020)	(0.025)	(0.015)	(0.023)	(0.023)	(0.021)	(0.008)	(0.030)	(0.017)
Q4	-0.036	0.012	-0.004	-0.046**	-0.004	0.020	0.021	0.001	0.020	-0.010	-0.012*	-0.019	0.017
	(0.032)	(0.038)	(0.036)	(0.017)	(0.028)	(0.029)	(0.016)	(0.024)	(0.021)	(0.019)	(0.006)	(0.024)	(0.015)
High saturation	-0.017	-0.003	-0.013	-0.006	-0.009	0.015	0.016	0.009	-0.011	0.006	-0.004	-0.039**	0.031***
	(0.018)	(0.027)	(0.021)	(0.009)	(0.018)	(0.018)	(0.009)	(0.013)	(0.012)	(0.008)	(0.003)	(0.016)	(0.011)
High saturation × Q3	-0.037	-0.005	0.036	0.026	-0.012	-0.024	-0.021	0.008	-0.019	-0.003	0.013	0.044	-0.032
	(0.109)	(0.049)	(0.038)	(0.017)	(0.028)	(0.034)	(0.018)	(0.026)	(0.027)	(0.016)	(0.008)	(0.043)	(0.025)
High saturation × Q4	0.046*	-0.020	-0.036	-0.005	0.052	0.011	-0.001	-0.013	-0.009	0.017	0.005	0.008	-0.020
	(0.023)	(0.047)	(0.046)	(0.017)	(0.036)	(0.034)	(0.017)	(0.028)	(0.021)	(0.026)	(0.009)	(0.029)	(0.016)
Observations	83	83	83	83	83	83	83	83	83	83	83	83	83
R^2	0.65	0.60	0.62	0.67	0.60	0.65	0.69	0.63	0.64	0.67	0.60	0.84	0.82
Low saturation outcome mean	0.49	0.21	0.27	0.23	0.16	0.09	0.04	0.27	0.29	0.38	0.06	0.88	0.07
Low saturation outcome std. dev.	0.10	0.08	0.07	0.03	0.05	0.05	0.03	0.04	0.04	0.04	0.01	0.08	0.05

Notes: Each specification is estimated using OLS, and includes randomization block fixed effects. All control municipalities and municipalities for which electoral data is unavailable are excluded. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Figure A2: The 128 municipalities included in this study



The regression estimates in panels A and C further show that the Facebook ads reached similar demographics of Facebook users in low and high saturation municipalities. Each coefficient represents the difference in a given share by user or ad consumption characteristic. While there are several statistically significant differences, the differences are small relative to the outcome means. Panels B and D further show that ads reporting different levels of irregularities did not systematically reach specific demographic subgroups. These results thus indicate that the high and low saturation ad campaigns reached similar types of audiences, and thus that differences in treatment effect by municipal saturation are unlikely to reflect differences in the types of voters who received the ad.

A.5 Relationship to pre-analysis plan

A.5.1 Administrative electoral outcomes

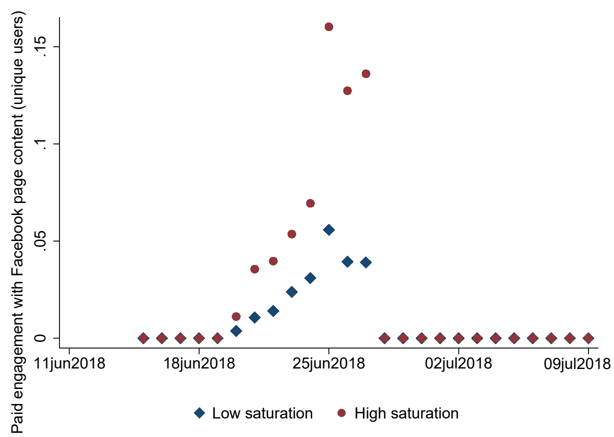
Our main analyses follow the specifications outlined in our pre-analysis plan (PAP; available at socialscienceregistry.org/trials/3135). At the precinct level:

1. Equation (1) corresponds to equation (5) in the PAP.

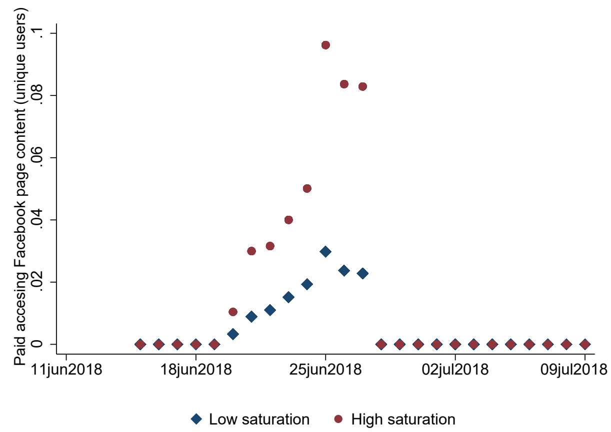
2. Equation (2) corresponds to equation (6) in the PAP, although equation (6) in the PAP is equivalently written in terms of interactions with municipal saturation instead of separate treatment conditions for each level of saturation.

Figure A3: Trends in engagement with the Facebook ads during the information campaign

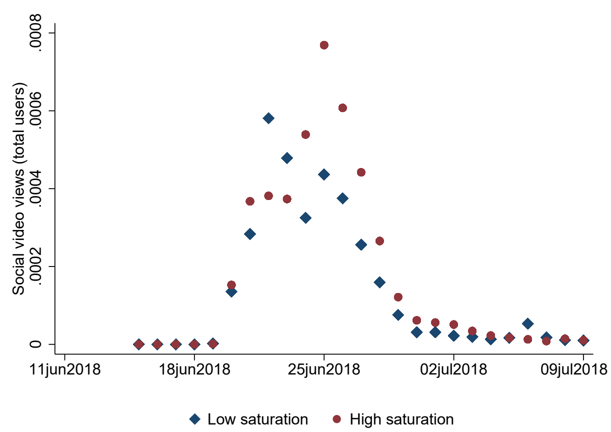
(a) Paid-for impressions



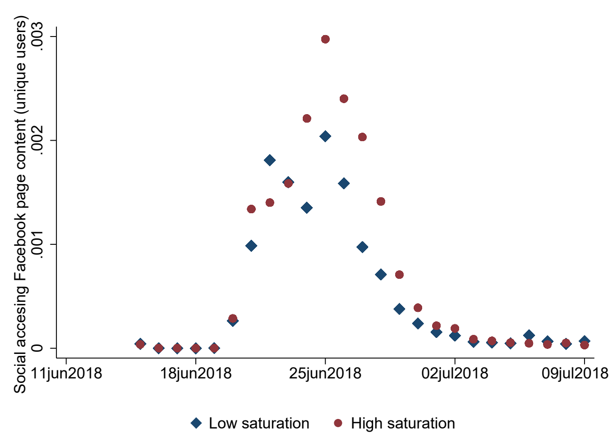
(b) Paid-for unique viewers



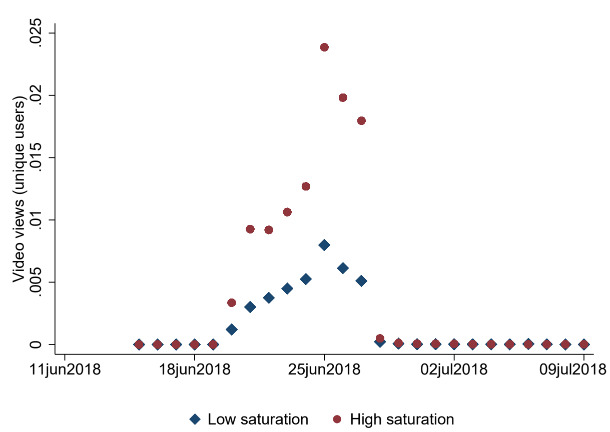
(c) Organic impressions



(d) Organic unique viewers



(e) Unique views (of at least 3 seconds)



(f) Unique views (of entire video)

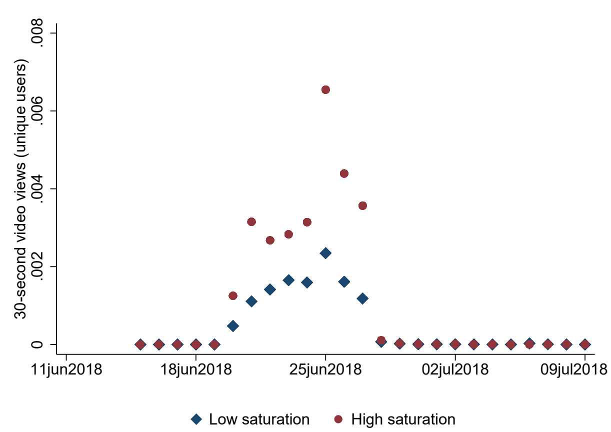
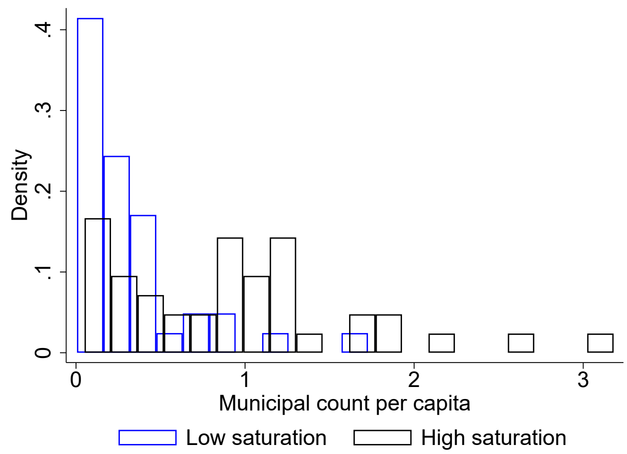
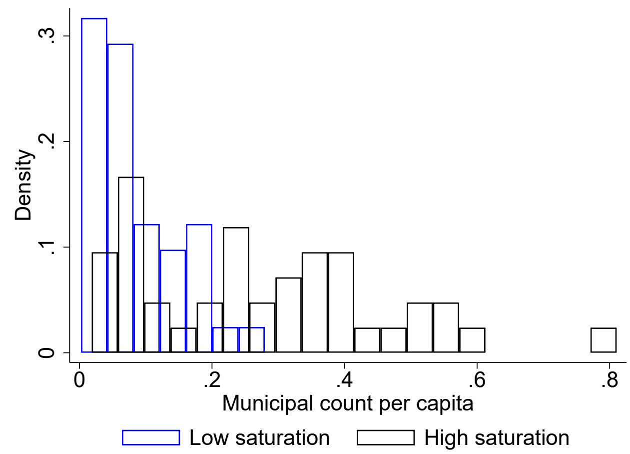


Figure A4: Distribution of engagement with the Facebook ads during the information campaign, by municipal saturation

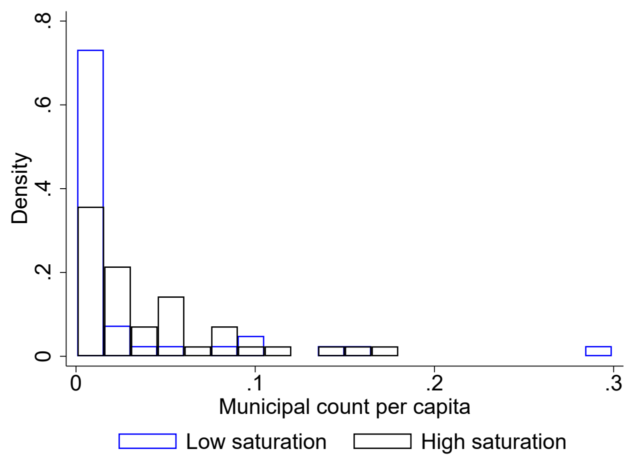
(a) Paid-for impressions



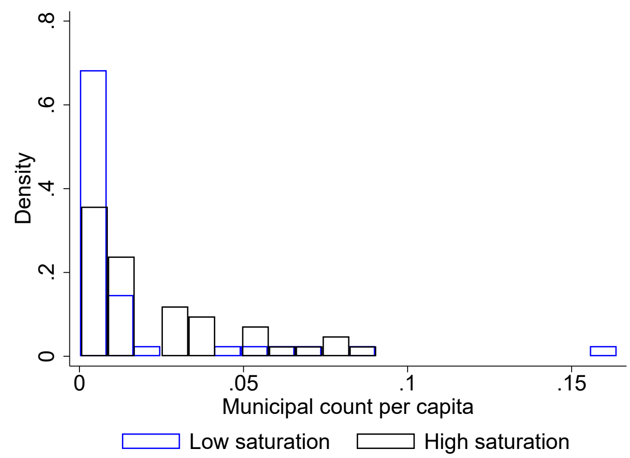
(b) Paid-for unique viewers



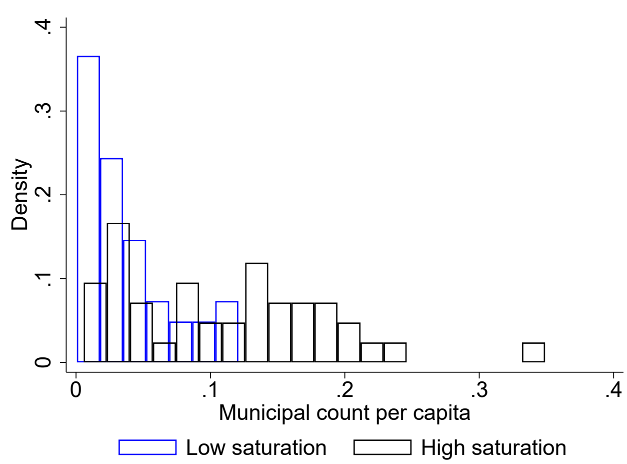
(c) Organic impressions



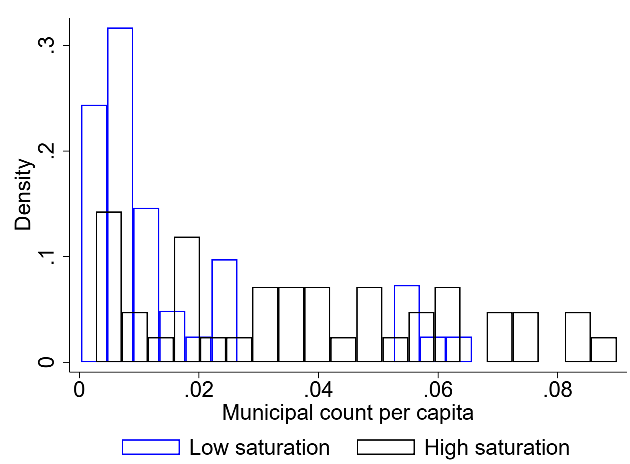
(d) Organic unique viewers



(e) Unique views (of at least 3 seconds)



(f) Unique views (of entire video)



3. Equation (3) is described in Section 4.2 of the PAP.

4. Equation (4) in the PAP corresponds to our tests between the basic and common knowledge variants of the treatment (see Section A.8).

At the municipal level, equation (4) corresponds to equation (1) in the PAP. Equations (2) and (3) in the PAP are less general than equation (1) in the PAP, and are thus excluded.

However, we also note one difference from the pre-specified approach and one clarification:

1. The one block of municipalities containing only two municipalities was excluded from the main analyses. We excluded this "rump" block (which contained two small municipalities comprising 23 precincts) on the basis that the treatment assignment probabilities vary from the other 42 blocks and that no electoral data was available for one segment within one of these municipalities. However, Table A3 reports similar results for our main estimates when these two municipalities are included; these regressions use inverse weights to adjust for the different probability of treatment assignment in the two-municipality block.

2. Although the PAP specified that we would examine heterogeneous effects by the share of irregularities reported, it did not specify how this important moderator capturing the content of the information provided would be operationalized. Our approach follows Cavalcanti, Daniele and Galletta (2018) and Larreguy, Marshall and Snyder (2020) in splitting the sample by municipalities that are above and below the median level of irregularities and by quartile of the municipal irregularities distribution.

As is usually the case, our PAP specified the main specifications. Accordingly, the mechanism analyses, extensions, and robustness checks were not prespecified.

A.5.2 Parallel panel survey

In addition to the precinct-level analysis of voting behavior that the main paper focuses on, we also conducted a small panel survey that yielded around 2,000 registered voters within our 128 municipalities. Participants were recruited on Facebook and via a landline telephone list. The 20-minute baseline survey was conducted over 3 weeks in early June 2018, while the 20-minute endline survey was conducted over the two months after the election. In addition to variation across municipalities in Facebook-level intervention, we further randomized (within blocks of five similar respondents) whether participants in treated municipalities and 23 of 43 control municipalities received a version of the Facebook ad via WhatsApp in the days before the election; respondents were treated with 80% probability. To view the ad, respondents needed to click to download the video because videos sent via WhatsApp only download automatically for existing contacts.

Table A3: Effect of Facebook ads on precinct-level municipal incumbent party vote share, including the block containing only two municipalities

	Incumbent party vote (share of turnout)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Facebook ads	0.023	0.060**	0.068**	0.056**	0.073***					
	(0.016)	(0.023)	(0.029)	(0.022)	(0.027)					
Facebook ads × Above-median irregularities		-0.087**	-0.076*							
		(0.034)	(0.042)							
Facebook ads × Irregularities Q3				-0.090**	-0.049					
				(0.044)	(0.045)					
Facebook ads × Irregularities Q4				-0.057	-0.086*					
				(0.038)	(0.045)					
Spillover	0.008	0.026	0.030	0.028	0.041					
	(0.016)	(0.022)	(0.027)	(0.021)	(0.025)					
Spillover × Above-median irregularities		-0.034	-0.042							
		(0.036)	(0.041)							
Spillover × Irregularities Q3				-0.003	-0.010					
				(0.039)	(0.041)					
Spillover × Irregularities Q4				-0.045	-0.066					
				(0.044)	(0.046)					
Facebook ads in high saturation						0.031*	0.079***	0.079**	0.073***	0.079**
						(0.018)	(0.027)	(0.034)	(0.026)	(0.033)
Facebook ads in high saturation × Above-median irregularities							-0.120***	-0.072		
							(0.040)	(0.056)		
Facebook ads in high saturation × Irregularities Q3									-0.135**	-0.072
									(0.052)	(0.058)
Facebook ads in high saturation × Irregularities Q4									-0.077*	-0.042
									(0.044)	(0.062)
Facebook ads in low saturation						0.009	0.022	0.009	0.024	0.028
						(0.018)	(0.027)	(0.031)	(0.026)	(0.031)
Facebook ads in low saturation × Above-median irregularities							-0.024	-0.030		
							(0.041)	(0.047)		
Facebook ads in low saturation × Irregularities Q3									0.003	-0.028
									(0.044)	(0.052)
Facebook ads in low saturation × Irregularities Q4									-0.038	-0.051
									(0.050)	(0.056)
Spillover in high saturation						0.029	0.078***	0.072**	0.072***	0.072**
						(0.019)	(0.028)	(0.035)	(0.026)	(0.034)
Spillover in high saturation × Above-median irregularities							-0.124***	-0.065		
							(0.041)	(0.058)		
Spillover in high saturation × Irregularities Q3									-0.141**	-0.068
									(0.055)	(0.060)
Spillover in high saturation × Irregularities Q4									-0.079*	-0.029
									(0.043)	(0.064)
Spillover in low saturation						0.007	0.009	-0.003	0.011	0.016
						(0.019)	(0.026)	(0.031)	(0.026)	(0.030)
Spillover in low saturation × Above-median irregularities							0.002	0.002		
							(0.041)	(0.048)		
Spillover in low saturation × Irregularities Q3									0.050	0.023
									(0.044)	(0.053)
Spillover in low saturation × Irregularities Q4									-0.028	-0.035
									(0.050)	(0.056)
Observations	13,277	13,277	13,277	13,277	13,277	13,254	13,254	13,254	13,254	13,254
R^2	0.51	0.53	0.60	0.56	0.63	0.53	0.57	0.66	0.59	0.68
Control outcome mean	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
Control outcome std. dev.	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
Test: null effect of Facebook ads below median (*p* value)		0.01	0.02	0.01	0.01					
Test: null effect of Facebook ads above median (*p* value)		0.23	0.74							
Test: null effect of Facebook ads in Q3 (*p* value)				0.33	0.49					
Test: null effect of Facebook ads in Q4 (*p* value)				0.95	0.63					
Test: null effect of spillover below median (*p* value)		0.24	0.26	0.19	0.11					
Test: null effect of spillover above median (*p* value)		0.77	0.62							
Test: null effect of spillover in Q3 (*p* value)				0.43	0.28					
Test: null effect of spillover in Q4 (*p* value)				0.64	0.43					
Test: same effect of Facebook ads in high and low (*p* value)						0.27				
Test: same effect of Facebook ads in high and low below median (*p* value)							0.03	0.01	0.06	0.04
Test: same effect of Facebook ads in high and low above median (*p* value)							0.01	0.42		
Test: same effect of Facebook ads in high and low in Q3 (*p* value)									0.03	0.89
Test: same effect of Facebook ads in high and low in Q4 (*p* value)									0.78	0.24
Test: same effect of spillovers in high and low (*p* value)						0.30				
Test: same effect of spillovers in high and low below median (*p* value)							0.01	0.01	0.02	0.03
Test: same effect of spillovers in high and low above median (*p* value)							0.00	0.21		
Test: same effect of spillovers in high and low in Q3 (*p* value)									0.02	0.75
Test: same effect of spillovers in high and low in Q4 (*p* value)									0.64	0.20
Treatment × covariate interactions			✓		✓			✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. The test statistics at the foot of the table report *p* values from two-sided hypothesis tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Our PAP specified a number of analyses of our survey data. However, we do not believe that the survey sample or the within-survey treatment condition are sufficiently comparable to the field experiment to illuminate the mechanisms driving electorate-level voting behavior. First, the sample is unrepresentative of the electorate in a number of respects. Baseline survey participants were more likely to come from larger municipalities and were more politically engaged, while only 10% of baseline respondents completed the endline survey. Second, the endline sample turned out to be quite imbalanced with respect to municipal saturation level, registering statistically significant differences across most predetermined municipal-level covariates (in contrast with the field experiment). This renders it difficult to reliably identify effects of the Facebook ad treatments within the panel survey sample. Moreover, we lack the location information required to assign survey respondents to segments within municipalities, and thus could not estimate direct and indirect effects. Third, while the integrity of the randomization of the WhatsApp-level treatment was maintained, this treatment condition differed in important ways from the intervention on Facebook: the ads delivered by WhatsApp required the respondent to approve the sender in order to have access to the video, which may have hindered engagement; unlike the naturalistic Facebook intervention, survey respondents were recruited for an academic study and told to expect further potential communication; and WhatsApp delivered the ads in private, whereas Facebook's public forum allowed others to engage with the ad and likely signaled a more extensive campaign. The ad delivered via WhatsApp is thus less naturalistic, accessible, and socially interactive than the main treatment condition on Facebook. For all these reasons, we decided not to include the parallel panel survey as part of this study.

Nevertheless, the unreported results provide tentative evidence to suggest that the ads' content was comprehensible, regarded as credible, and could have influenced voting behavior. We find that respondents treated with the WhatsApp message were significantly more likely to report having received the WhatsApp message during the endline survey conducted several weeks after the election. Moreover, respondents who recalled the ad were more significantly likely to accurately identify the level of irregularities in their municipality. Neither response is necessary for the ad to produce effects, but these findings suggest that viewers internalized the content of Borde Político's ads. Descriptive data further suggests that a majority of those that recalled receiving the message regarded the ad's content as somewhat or very credible. Finally, we find very tentative evidence to suggest that self-reported votes for the incumbent increased among respondents that received the WhatsApp treatment in municipalities where government engaged in fewest irregularities. Full survey results are available upon request.

A.6 Balance tests

Table A4 reports balance tests based on the municipal-level treatment assignments in equation (4). We aggregate INEGI's microdata (for more than 22 million Mexicans) from the 2015 intercensal survey at the municipal level for many variables. For each predetermined variable, we report the p value associated with an F test of the restriction that $\beta_1 = \beta_2 = 0$. Only 4 of the 52 tests show a statistically significant difference at the $p < 0.1$ level.

Table A5 reports balance tests based on the segment-level treatment assignments in equation (2). For each predetermined variable, we report the p value associated with an F test of the restriction that $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 0$. Only 2 of 63 tests show a statistically significant difference at the $p < 0.1$ level. Because the 2015 intercensal survey in Mexico does not provide a geographic location for localities containing fewer than 50,000 people (and thus link precincts to intercensal microdata), we conduct balance tests using 2010 census data for the segment-level balance tests.

Given our focus on conditional average treatment effects with respect to ad content, we also conduct balance tests to ensure that balance continues to hold above and below the median of the irregularities distribution as well as across quartiles of the distribution using the exact specifications used for estimation. Accordingly, we examine differences by treatment condition within each bin by estimating interactive versions of equation (2). For each predetermined covariate, this entails conducting an F test of the restriction $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 0$ below the irregularities median/within Q1/2; for municipalities above the median level of irregularities or in Q3 and Q4 of the irregularities distribution, this entails testing whether the sum of the base and interactive coefficients is equal to 0 for each treatment condition separately; in addition, we test the null hypothesis that there are no differences between any treatment condition $\times$ median/quartile cell and the pure control group. The results in Tables A6 and A7 show that we do not observe relatively few imbalances above and below the median or within any quartile when using our main estimating equation.

Finally, we report summary statistics by quartile of the irregularities distribution in Table A8. Since reported irregularities were not randomized, there is no reason to expect balance across quartiles. Nevertheless, the table and associated tests suggest that Q1/2, Q3, and Q4 are relatively similar in terms of observable characteristics.

A.7 Alternative preregistered regression specifications

While we prefer the estimation strategies used in the main article, the pre-analysis plan also specified several additional approaches to estimation. Focusing on the main results in Tables 2 and 3, these decisions do not affect our findings. Indeed, our principal results are robust to excluding the lagged outcome (see Table A9), weighting precincts by the number of registered voters (see Table A10), and weighting municipalities with different numbers of segments equally (see Table A11).

Table A4: Municipal-level Facebook saturation treatment condition balance tests

Covariate	Control mean	Control std. dev.	F test (p value)	Covariate	Control mean	Control std. dev.	F test (p value)
Year of audit	2,015.85	0.35	0.23	Share with at least complete primary schooling in 2015	0.90	0.07	0.78
FISM resources received	101,887.58	130,758.39	0.21	Share with at least complete secondary schooling in 2015	0.58	0.15	0.61
FISM resources used	90,442.80	114,967.95	0.16	Share with higher education in 2015	0.29	0.14	0.54
FISM resources audited	76,844.26	89,970.69	0.26	Share disabled in 2015	0.02	0.01	0.34
Share of resources subject to irregularities	0.09	0.18	0.16	Share economically active in 2015	0.45	0.09	0.18
Municipal adult population in 2015	135,401.00	199,014.69	0.82	Share without health care in 2015	0.14	0.06	0.34
Registered voters in 2015	2,277.41	2,470.68	0.71	Share with state workers healthcare in 2015	0.01	0.01	0.13
Municipal incumbent seeking re-election	0.29	0.46	0.03	Average number of occupants per dwelling in 2015	4.02	0.42	0.68
Municipal incumbent vote share in 2015	0.42	0.11	0.81	Average number of occupants per room in 2015	1.51	0.40	0.21
Municipal turnout in 2015	0.60	0.14	0.90	Share of households with more than 2 bedrooms in 2015	0.59	0.13	0.15
Federal MORENA-PT-PES vote share in 2015	0.10	0.08	0.92	Share of households with more than 3 rooms in 2015	0.66	0.15	0.22
Federal PAN-PRD-MC vote share in 2015	0.42	0.15	0.44	Share of households with non-dirt floor in 2015	0.92	0.09	0.38
Federal PRI-PVEM-NA vote share in 2015	0.43	0.17	0.22	Share of households with a toilet in 2015	0.93	0.10	0.04
Federal turnout in 2015	0.56	0.14	0.86	Share of households with water in 2015	0.92	0.08	0.13
Share women in 2015	0.51	0.01	0.27	Share of households with drainage in 2015	0.85	0.15	0.94
Share working age in 2015	0.61	0.05	0.40	Share of households with electricity in 2015	0.98	0.04	0.68
Share over 65 in 2015	0.07	0.02	0.86	Share of households with water, drainage, and electricity in 2015	0.80	0.17	0.91
Share married in 2015	0.56	0.03	0.25	Share of households with a washing machine in 2015	0.51	0.25	0.16
Average children per woman in 2015	2.58	0.40	0.90	Share of households with a landline telephone in 2015	0.22	0.16	0.25
Share of households with a male head in 2015	0.27	0.05	0.41	Share of households with a radio in 2015	0.63	0.16	0.26
Share indigenous speakers in 2015	0.21	0.31	0.04	Share of households with a fridge in 2015	0.70	0.24	0.17
Average years of schooling in 2015	6.50	1.39	0.58	Share of households with a cell phone in 2015	0.65	0.23	0.60
Average years of schooling for women in 2015	6.44	1.43	0.46	Share of households with a television in 2015	0.84	0.18	0.09
Average years of schooling for men in 2015	6.56	1.37	0.68	Share of households with a car in 2015	0.31	0.18	0.14
Share illiterate in 2015	0.11	0.09	0.46	Share of households with a computer in 2015	0.20	0.15	0.57
Share with no schooling in 2015	0.10	0.07	0.80	Share of households with internet in 2015	0.19	0.16	0.52

Notes: The F tests report the p value associated with the test of the restriction that $\beta_1 = \beta_2 = 0$ in equation (4) in the main paper. Each specification is estimated using OLS, and includes municipal saturation conditions and randomization block fixed effects. All observations are weighted equally by municipality.

Table A5: Segment-level Facebook ad treatment condition balance tests

Covariate	Control mean	Control std. dev.	F test (p value)	Covariate	Control mean	Control std. dev.	F test (p value)
Year of audit	2,015.87	0.33	0.39	Average years of schooling for men in 2010	7.89	2.42	0.87
FISM resources received	109,254.87	134,275.10	0.39	Share illiterate in 2010	0.12	0.11	0.18
FISM resources used	95,119.56	120,793.89	0.13	Share with no schooling in 2010	0.11	0.09	0.28
FISM resources audited	81,344.42	89,436.74	0.29	Share with at least incomplete primary schooling in 2010	0.89	0.09	0.28
Share of resources subject to irregularities	0.08	0.16	0.35	Share with at least complete primary schooling in 2010	0.73	0.16	0.45
Municipal adult population in 2010	209,177.31	274,260.96	0.88	Share with at least incomplete secondary schooling in 2010	0.55	0.21	0.77
Segment adult population in 2010	23,131.00	20,766.99	0.57	Share with at least complete secondary schooling in 2010	0.50	0.21	0.75
Precinct adult population in 2010	2,034.50	2,329.59	0.74	Share with higher education in 2010	0.28	0.21	0.85
Registered voters in 2015	2,332.11	2,579.79	0.79	Share disabled in 2010	0.04	0.02	0.70
Municipal incumbent seeking re-election	0.29	0.45	0.20	Share economically active in 2010	0.37	0.07	0.19
Municipal incumbent vote share in 2015	0.43	0.11	0.59	Share without health care in 2010	0.34	0.17	0.78
Municipal turnout in 2015	0.59	0.14	0.61	Share with state workers healthcare in 2010	0.04	0.05	0.29
Federal MORENA-PT-PES vote share in 2015	0.10	0.08	0.58	Average number of occupants per dwelling in 2010	4.24	0.63	0.30
Federal PAN-PRD-MC vote share in 2015	0.42	0.15	0.59	Average number of occupants per room in 2010	1.33	0.45	0.09
Federal PRI-PVEM-NA vote share in 2015	0.42	0.16	0.61	Share of households with more than 2 bedrooms in 2010	0.59	0.17	0.21
Federal turnout in 2015	0.55	0.14	0.45	Share of households with more than 3 rooms in 2010	0.68	0.20	0.12
Number of households in 2010	818.18	1,022.19	0.83	Share of households with non-dirt floor in 2010	0.90	0.12	0.49
Private dwellings in 2010	1,047.57	1,493.27	0.93	Share of households with a toilet in 2010	0.90	0.17	0.13
Population in 2010	3,415.12	3,906.55	0.79	Share of households with water in 2010	0.85	0.22	0.12
Share women in 2010	0.51	0.02	0.53	Share of households with drainage in 2010	0.82	0.25	0.21
Share working age in 2010	0.61	0.06	0.43	Share of households with electricity in 2010	0.96	0.08	0.28
Share over 65 in 2010	0.06	0.03	0.97	Share of households with water, drainage, and electricity in 2010	0.73	0.30	0.46
Share married in 2010	0.55	0.05	0.77	Share of households with a washing machine in 2010	0.51	0.30	0.13
Average children per woman in 2010	2.59	0.58	0.68	Share of households with a landline telephone in 2010	0.33	0.25	0.78
Share of households with a male head in 2010	0.77	0.08	0.32	Share of households with a radio in 2010	0.72	0.16	0.21
Share born out of state in 2010	0.12	0.14	0.32	Share of households with a fridge in 2010	0.68	0.29	0.14
Share Catholic in 2010	0.84	0.16	0.43	Share of households with a cell phone in 2010	0.51	0.30	0.17
Share non-Catholic in 2010	0.10	0.12	0.27	Share of households with a television in 2010	0.84	0.20	0.14
Share without religion in 2010	0.03	0.04	1.00	Share of households with a car in 2010	0.34	0.23	0.19
Share indigenous speakers in 2010	0.18	0.33	0.05	Share of households with a computer in 2010	0.21	0.21	0.88
Average years of schooling in 2010	7.67	2.38	0.89	Share of households with internet in 2010	0.15	0.18	0.91
Average years of schooling for women in 2010	7.47	2.38	0.77				

Notes: The F tests report the p value associated with the test of the restriction that $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 0$ in equation (1) in the main paper. Each specification is estimated using OLS, and includes segment-level direct and indirect treatment conditions (by saturation) and randomization block fixed effects. All segments are weighted equally and by the inverse probability of treatment assignment.

Table A6: Segment-level Facebook ad treatment condition balance tests, above and below the median of the municipal irregularities distribution

Covariate	Below-median F test (p value)	Above-median F test (p value)	Above- and below-median F test (p value)	Covariate	Below-median F test (p value)	Above-median F test (p value)	Above- and below-median F test (p value)
Year of audit	0.089	0.968	0.408	Average years of schooling for men in 2010	0.397	0.622	0.608
FISM resources received	0.483	0.394	0.545	Share illiterate in 2010	0.106	0.733	0.290
FISM resources used	0.110	0.329	0.147	Share with no schooling in 2010	0.057	0.625	0.160
FISM resources audited	0.189	0.512	0.400	Share with at least incomplete primary schooling in 2010	0.057	0.625	0.160
Share of resources subject to irregularities	0.078	0.949	0.277	Share with at least complete primary schooling in 2010	0.061	0.722	0.218
Municipal adult population in 2010	0.713	0.795	0.900	Share with at least incomplete secondary schooling in 2010	0.294	0.571	0.500
Segment adult population in 2010	0.629	0.785	0.828	Share with at least complete secondary schooling in 2010	0.236	0.532	0.420
Precinct adult population in 2010	0.172	0.251	0.213	Share with higher education in 2010	0.464	0.445	0.546
Registered voters in 2015	0.230	0.198	0.193	Share disabled in 2010	0.694	0.244	0.483
Municipal incumbent seeking re-election	0.721	0.325	0.482	Share economically active in 2010	0.065	0.972	0.317
Municipal turnout in 2015	0.748	0.649	0.832	Share without health care in 2010	0.414	0.008	0.041
Federal MORENA-PT-PES vote share in 2015	0.015	0.032	0.016	Share with state workers healthcare in 2010	0.173	0.038	0.051
Federal PAN-PRD-MC vote share in 2015	0.242	0.746	0.501	Average number of occupants per dwelling in 2010	0.390	0.642	0.560
Federal PRI-PVEM-NA vote share in 2015	0.955	0.180	0.581	Average number of occupants per room in 2010	0.303	0.360	0.305
Federal turnout in 2015	0.532	0.765	0.769	Share of households with more than 2 bedrooms in 2010	0.625	0.509	0.584
Number of households in 2010	0.194	0.276	0.230	Share of households with more than 3 rooms in 2010	0.408	0.560	0.458
Private dwellings in 2010	0.301	0.289	0.300	Share of households with non-dirt floor in 2010	0.294	0.442	0.328
Population in 2010	0.213	0.168	0.186	Share of households with a toilet in 2010	0.241	0.436	0.437
Share women in 2010	0.965	0.016	0.122	Share of households with water in 2010	0.425	0.000	0.000
Share working age in 2010	0.136	0.751	0.352	Share of households with drainage in 2010	0.273	0.414	0.319
Share over 65 in 2010	0.452	0.158	0.235	Share of households with electricity in 2010	0.237	0.911	0.468
Share married in 2010	0.992	0.078	0.369	Share of households with water, drainage, and electricity in 2010	0.346	0.768	0.579
Average children per woman in 2010	0.194	0.372	0.253	Share of households with a washing machine in 2010	0.240	0.493	0.378
Share of households with a male head in 2010	0.379	0.049	0.080	Share of households with a landline telephone in 2010	0.740	0.737	0.844
Share born out of state in 2010	0.067	0.524	0.173	Share of households with a radio in 2010	0.154	0.705	0.310
Share Catholic in 2010	0.491	0.077	0.142	Share of households with a fridge in 2010	0.268	0.625	0.424
Share non-Catholic in 2010	0.718	0.108	0.229	Share of households with a cell phone in 2010	0.092	0.921	0.350
Share without religion in 2010	0.995	0.935	0.997	Share of households with a television in 2010	0.180	0.348	0.205
Share indigenous speakers in 2010	0.168	0.213	0.151	Share of households with a car in 2010	0.401	0.645	0.563
Average years of schooling in 2010	0.505	0.496	0.603	Share of households with a computer in 2010	0.792	0.764	0.880
Average years of schooling for women in 2010	0.277	0.581	0.468	Share of households with internet in 2010	0.880	0.828	0.937

Notes: The F tests report the p value associated with a test of the null hypothesis of balance across all treatment conditions within each quartile (or across all treatment condition $\times$ quartile and the control group), based on estimating the interactive version of equation (2) in the main paper. Each specification is estimated using OLS, and includes segment-level direct and indirect treatment conditions, their interaction with quartile fixed effects (including the lower-order quartile terms), municipal incumbent vote share in 2015, and randomization block fixed effects. Due to the inclusion of the lagged vote share outcome in the main specification, we cannot conduct balance tests for this covariate. All segments are weighted equally and by the inverse probability of treatment assignment.

Table A7: Segment-level Facebook ad treatment condition balance tests, by quartile of the municipal irregularities distribution

Covariate	Q1/2 F test (p value)	Q3 F test (p value)	Q4 F test (p value)	All Q F test (p value)	Covariate	Q1/2 F test (p value)	Q3 F test (p value)	Q4 F test (p value)	All Q F test (p value)
Year of audit	0.10	0.71	0.53	0.52	Average years of schooling for men in 2010	0.37	0.40	0.63	0.51
FISM resources received	0.45	0.50	0.37	0.59	Share illiterate in 2010	0.10	0.82	0.16	0.19
FISM resources used	0.10	0.54	0.31	0.22	Share with no schooling in 2010	0.05	0.91	0.18	0.17
FISM resources audited	0.18	0.54	0.26	0.38	Share with at least incomplete primary schooling in 2010	0.05	0.91	0.18	0.17
Share of resources subject to irregularities	0.38	0.34	0.26	0.53	Share with at least complete primary schooling in 2010	0.05	0.82	0.51	0.30
Municipal adult population in 2010	0.67	0.93	0.49	0.89	Share with at least incomplete secondary schooling in 2010	0.27	0.39	0.57	0.46
Segment adult population in 2010	0.66	0.46	0.77	0.82	Share with at least complete secondary schooling in 2010	0.21	0.33	0.58	0.36
Precinct adult population in 2010	0.16	0.28	0.21	0.19	Share with higher education in 2010	0.46	0.20	0.55	0.39
Registered voters in 2015	0.21	0.28	0.15	0.19	Share disabled in 2010	0.68	0.29	0.43	0.51
Municipal incumbent seeking re-election	0.75	0.22	0.96	0.62	Share economically active in 2010	0.07	0.57	0.65	0.33
Municipal turnout in 2015	0.74	0.32	0.88	0.77	Share without health care in 2010	0.35	0.00	0.58	0.00
Federal MORENA-PT-PES vote share in 2015	0.01	0.00	0.14	0.00	Share with state workers healthcare in 2010	0.23	0.34	0.00	0.02
Federal PAN-PRD-MC vote share in 2015	0.23	0.26	0.66	0.40	Average number of occupants per dwelling in 2010	0.36	0.90	0.60	0.75
Federal PRI-PVEM-NA vote share in 2015	0.94	0.02	0.46	0.15	Average number of occupants per room in 2010	0.27	0.60	0.29	0.28
Federal turnout in 2015	0.52	0.37	0.98	0.80	Share of households with more than 2 bedrooms in 2010	0.61	0.11	0.37	0.17
Number of households in 2010	0.17	0.24	0.23	0.19	Share of households with more than 3 rooms in 2010	0.39	0.51	0.64	0.48
Private dwellings in 2010	0.26	0.09	0.23	0.13	Share of households with non-dirt floor in 2010	0.29	0.06	0.02	0.03
Population in 2010	0.19	0.22	0.17	0.16	Share of households with a toilet in 2010	0.40	0.02	0.10	0.06
Share women in 2010	0.98	0.28	0.04	0.11	Share of households with water in 2010	0.41	0.00	0.00	0.00
Share working age in 2010	0.12	0.35	0.40	0.18	Share of households with drainage in 2010	0.24	0.93	0.29	0.51
Share over 65 in 2010	0.45	0.56	0.00	0.02	Share of households with electricity in 2010	0.36	0.35	0.52	0.49
Share married in 2010	0.99	0.26	0.50	0.64	Share of households with water, drainage, and electricity in 2010	0.35	0.77	0.75	0.73
Average children per woman in 2010	0.18	0.17	0.32	0.16	Share of households with a washing machine in 2010	0.25	0.57	0.35	0.39
Share of households with a male head in 2010	0.44	0.50	0.19	0.23	Share of households with a landline telephone in 2010	0.71	0.62	0.70	0.85
Share born out of state in 2010	0.06	0.11	0.86	0.13	Share of households with a radio in 2010	0.15	0.88	0.38	0.42
Share Catholic in 2010	0.50	0.15	0.56	0.33	Share of households with a fridge in 2010	0.27	0.64	0.12	0.26
Share non-Catholic in 2010	0.73	0.21	0.27	0.33	Share of households with a cell phone in 2010	0.10	0.86	0.40	0.34
Share without religion in 2010	1.00	0.29	0.22	0.44	Share of households with a television in 2010	0.14	0.93	0.07	0.15
Share indigenous speakers in 2010	0.17	0.78	0.23	0.28	Share of households with a car in 2010	0.41	0.52	0.43	0.49
Average years of schooling in 2010	0.49	0.46	0.58	0.60	Share of households with a computer in 2010	0.79	0.44	0.69	0.79
Average years of schooling for women in 2010	0.27	0.46	0.55	0.45	Share of households with internet in 2010	0.87	0.47	0.76	0.86

Notes: The F tests report the p value associated with a test of the null hypothesis of balance across all treatment conditions within each quartile (or across all treatment condition $\times$ quartile and the control group), based on estimating the interactive version of equation (2) in the main paper. Each specification is estimated using OLS, and includes segment-level direct and indirect treatment conditions, their interaction with quartile fixed effects (including the lower-order quartile terms), municipal incumbent vote share in 2015, and randomization block fixed effects. Due to the inclusion of the lagged vote share outcome in the main specification, we cannot conduct balance tests for this covariate. All segments are weighted equally and by the inverse probability of treatment assignment.

Table A8: Covariate summary statistics, by irregularities quartile

Covariate	Q1/2 mean	Q1/2 std. dev.	Q3 mean	Q3 std. dev.	Q4 mean	Q4 std. dev.	F test (p value)	Covariate	Q1/2 mean	Q1/2 std. dev.	Q3 mean	Q3 std. dev.	Q4 mean	Q4 std. dev.	F test (p value)
Year of audit	2,015.79	0.41	2,015.81	0.40	2,015.90	0.30	0.13	Share with at least complete primary schooling in 2015	0.90	0.07	0.89	0.06	0.90	0.08	0.36
FISM resources received	89,571.24	80,891.08	97,398.24	155,292.87	69,799.79	64,174.10	0.01	Share with at least complete secondary schooling in 2015	0.58	0.15	0.57	0.14	0.61	0.17	0.24
FISM resources used	78,278.60	75,489.04	84,407.14	135,727.98	64,538.35	63,542.13	0.02	Share with higher education in 2015	0.29	0.14	0.29	0.13	0.33	0.15	0.62
FISM resources audited	71,427.53	63,777.46	73,077.46	106,016.37	57,425.57	55,601.81	0.12	Share disabled in 2015	0.02	0.01	0.02	0.01	0.02	0.01	0.02
Share of resources subject to irregularities	0.00	0.00	0.02	0.02	0.32	0.26	0.00	Share economically active in 2015	0.46	0.08	0.45	0.08	0.47	0.09	0.12
Municipal adult population in 2010	187,494.81	285,606.69	111,768.58	166,090.28	99,498.10	118,902.93	0.07	Share without health care in 2015	0.15	0.07	0.13	0.05	0.14	0.06	0.25
Registered voters in 2015	2,416.12	4,078.05	2,272.64	2,564.53	2,373.82	3,244.35	0.80	Share with state workers healthcare in 2015	0.01	0.01	0.00	0.00	0.01	0.02	0.59
Municipal incumbent seeking re-election	0.32	0.47	0.26	0.44	0.26	0.44	0.66	Average number of occupants per dwelling in 2015	4.05	0.40	3.99	0.44	3.90	0.37	0.76
Municipal incumbent vote share in 2015	0.42	0.13	0.42	0.13	0.44	0.13	0.14	Average number of occupants per room in 2015	1.46	0.29	1.47	0.35	1.48	0.40	0.98
Municipal turnout in 2015	0.60	0.14	0.63	0.14	0.58	0.13	0.13	Share of households with more than 2 bedrooms in 2015	0.61	0.11	0.59	0.12	0.60	0.11	0.52
Federal MORENA-PT-PES vote share in 2015	0.10	0.09	0.08	0.08	0.08	0.07	0.02	Share of households with more than 3 rooms in 2015	0.69	0.12	0.67	0.14	0.67	0.15	0.89
Federal PAN-PRD-MC vote share in 2015	0.41	0.14	0.38	0.17	0.43	0.15	0.64	Share of households with non-dirt floor in 2015	0.93	0.07	0.93	0.07	0.91	0.10	0.31
Federal PRI-PVEM-NA vote share in 2015	0.43	0.15	0.47	0.17	0.42	0.17	0.09	Share of households with a toilet in 2015	0.96	0.04	0.94	0.10	0.93	0.09	0.15
Federal turnout in 2015	0.56	0.13	0.59	0.14	0.55	0.14	0.26	Share of households with water in 2015	0.90	0.10	0.92	0.08	0.89	0.11	0.19
Share women in 2015	0.51	0.01	0.51	0.01	0.51	0.01	0.32	Share of households with drainage in 2015	0.86	0.16	0.87	0.14	0.82	0.19	0.62
Share working age in 2015	0.62	0.04	0.62	0.04	0.62	0.05	0.42	Share of households with electricity in 2015	0.98	0.02	0.98	0.02	0.97	0.04	0.31
Share over 65 in 2015	0.07	0.02	0.07	0.03	0.07	0.02	0.06	Share of households with water, drainage, and electricity in 2015	0.80	0.18	0.81	0.15	0.76	0.21	0.82
Share married in 2015	0.56	0.03	0.56	0.03	0.56	0.02	0.61	Share of households with a washing machine in 2015	0.54	0.24	0.52	0.25	0.52	0.24	0.24
Average children per woman in 2015	2.55	0.40	2.59	0.37	2.56	0.46	0.24	Share of households with a landline telephone in 2015	0.23	0.17	0.25	0.16	0.24	0.14	0.97
Share of households with a male head in 2015	0.27	0.04	0.27	0.07	0.28	0.04	0.86	Share of households with a radio in 2015	0.65	0.15	0.65	0.13	0.64	0.15	0.89
Share indigenous speakers in 2015	0.15	0.24	0.23	0.34	0.15	0.27	0.02	Share of households with a fridge in 2015	0.69	0.23	0.70	0.24	0.74	0.24	0.36
Average years of schooling in 2015	6.53	1.39	6.47	1.35	6.86	1.46	0.62	Share of households with a cell phone in 2015	0.67	0.20	0.62	0.24	0.68	0.24	0.22
Average years of schooling for women in 2015	6.48	1.42	6.41	1.41	6.87	1.49	0.59	Share of households with a television in 2015	0.86	0.14	0.85	0.15	0.85	0.18	0.24
Average years of schooling for men in 2015	6.58	1.37	6.53	1.31	6.85	1.44	0.63	Share of households with a car in 2015	0.32	0.16	0.31	0.18	0.36	0.19	0.45
Share illiterate in 2015	0.10	0.08	0.12	0.08	0.10	0.09	0.17	Share of households with a computer in 2015	0.20	0.14	0.21	0.15	0.24	0.15	0.75
Share with no schooling in 2015	0.10	0.07	0.11	0.06	0.09	0.08	0.35	Share of households with internet in 2015	0.19	0.16	0.20	0.16	0.23	0.16	0.68

Notes: The F tests report the p value associated with a test of the null hypothesis of balance across irreguarlities quartile, based on estimating an OLS regression that includes randomization block fixed effects. All municipalities are weighted equally.

Table A9: Effect of Facebook ads on precinct-level municipal incumbent party vote share, excluding the lagged dependent variable

	Incumbent party vote (share of turnout)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Facebook ads	0.025	0.058**	0.070**	0.054**	0.075***					
	(0.017)	(0.026)	(0.029)	(0.026)	(0.028)					
Facebook ads × Above-median irregularities		-0.078**	-0.071*							
		(0.039)	(0.041)							
Facebook ads × Irregularities Q3				-0.087*	-0.046					
				(0.052)	(0.045)					
Facebook ads × Irregularities Q4				-0.044	-0.080*					
				(0.041)	(0.045)					
Spillover	0.009	0.018	0.029	0.020	0.039					
	(0.018)	(0.025)	(0.026)	(0.024)	(0.025)					
Spillover × Above-median irregularities		-0.013	-0.031							
		(0.040)	(0.040)							
Spillover × Irregularities Q3				0.013	0.002					
				(0.048)	(0.041)					
Spillover × Irregularities Q4				-0.022	-0.054					
				(0.044)	(0.046)					
Facebook ads in high saturation						0.028	0.077**	0.077**	0.071**	0.077**
						(0.020)	(0.031)	(0.034)	(0.030)	(0.034)
Facebook ads in high saturation × Above-median irregularities							-0.119***	-0.064		
							(0.044)	(0.056)		
Facebook ads in high saturation × Irregularities Q3									-0.142**	-0.065
									(0.059)	(0.058)
Facebook ads in high saturation × Irregularities Q4									-0.068	-0.032
									(0.048)	(0.063)
Facebook ads in low saturation						0.013	0.020	0.019	0.023	0.038
						(0.019)	(0.030)	(0.031)	(0.029)	(0.031)
Facebook ads in low saturation × Above-median irregularities							-0.010	-0.033		
							(0.043)	(0.045)		
Facebook ads in low saturation × Irregularities Q3									0.013	-0.038
									(0.048)	(0.053)
Facebook ads in low saturation × Irregularities Q4									-0.019	-0.047
									(0.051)	(0.055)
Spillover in high saturation						0.025	0.073**	0.067*	0.068**	0.068*
						(0.020)	(0.031)	(0.036)	(0.030)	(0.035)
Spillover in high saturation × Above-median irregularities							-0.117**	-0.052		
							(0.046)	(0.058)		
Spillover in high saturation × Irregularities Q3									-0.142**	-0.056
									(0.063)	(0.059)
Spillover in high saturation × Irregularities Q4									-0.063	-0.015
									(0.045)	(0.065)
Spillover in low saturation						0.007	-0.001	-0.004	0.002	0.015
						(0.021)	(0.030)	(0.030)	(0.029)	(0.030)
Spillover in low saturation × Above-median irregularities							0.024	0.012		
							(0.045)	(0.046)		
Spillover in low saturation × Irregularities Q3									0.066	0.028
									(0.051)	(0.052)
Spillover in low saturation × Irregularities Q4									-0.000	-0.019
									(0.052)	(0.056)
Observations	13,281	13,281	13,281	13,281	13,281	13,281	13,281	13,281	13,281	13,281
R^2	0.44	0.45	0.56	0.48	0.59	0.47	0.50	0.62	0.53	0.64
Control outcome mean	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
Control outcome std. dev.	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
Test: null effect of Facebook ads below median (*p* value)		0.03	0.02	0.04	0.01					
Test: null effect of Facebook ads above median (*p* value)		0.43	0.97							
Test: null effect of Facebook ads in Q3 (*p* value)				0.42	0.42					
Test: null effect of Facebook ads in Q4 (*p* value)				0.77	0.84					
Test: null effect of spillover below median (*p* value)		0.48	0.28	0.42	0.13					
Test: null effect of spillover above median (*p* value)		0.86	0.93							
Test: null effect of spillover in Q3 (*p* value)				0.39	0.17					
Test: null effect of spillover in Q4 (*p* value)				0.96	0.62					
Test: same effect of Facebook ads in high and low (*p* value)						0.45				
Test: same effect of Facebook ads in high and low below median (*p* value)							0.03	0.04	0.06	0.11
Test: same effect of Facebook ads in high and low above median (*p* value)							0.01	0.55		
Test: same effect of Facebook ads in high and low in Q3 (*p* value)									0.01	0.81
Test: same effect of Facebook ads in high and low in Q4 (*p* value)									1.00	0.29
Test: same effect of spillovers in high and low (*p* value)						0.39				
Test: same effect of spillovers in high and low below median (*p* value)							0.00	0.01	0.01	0.04
Test: same effect of spillovers in high and low above median (*p* value)							0.00	0.23		
Test: same effect of spillovers in high and low in Q3 (*p* value)									0.01	0.79
Test: same effect of spillovers in high and low in Q4 (*p* value)									0.92	0.23
Treatment × covariate interactions			✓		✓			✓		✓

Notes: Each specification is estimated using OLS, and includes randomization block fixed effects. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. The test statistics at the foot of the table report p values from two-sided hypothesis tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Table A10: Effect of Facebook ads on precinct-level municipal incumbent party vote share, weighting by registered voters

	Incumbent party vote (share of turnout)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Facebook ads	0.049**	0.061**	0.085***	0.092***	0.091***					
	(0.020)	(0.023)	(0.029)	(0.025)	(0.029)					
Facebook ads × Above-median irregularities		-0.085**	-0.127***							
		(0.034)	(0.037)							
Facebook ads × Irregularities Q3				-0.142***	-0.139***					
				(0.037)	(0.041)					
Facebook ads × Irregularities Q4				-0.078**	-0.105**					
				(0.034)	(0.047)					
Spillover	0.011	0.027	0.052*	0.040	0.060**					
	(0.019)	(0.022)	(0.028)	(0.024)	(0.028)					
Spillover × Above-median irregularities		-0.031	-0.098**							
		(0.036)	(0.039)							
Spillover × Irregularities Q3				-0.049	-0.099**					
				(0.041)	(0.045)					
Spillover × Irregularities Q4				-0.033	-0.093**					
				(0.035)	(0.047)					
Facebook ads in high saturation						0.070***	0.124***	0.122***	0.119***	0.133***
						(0.021)	(0.022)	(0.025)	(0.023)	(0.026)
Facebook ads in high saturation × Above-median irregularities							-0.161***	-0.139***		
							(0.031)	(0.037)		
Facebook ads in high saturation × Irregularities Q3									-0.197***	-0.151***
									(0.034)	(0.040)
Facebook ads in high saturation × Irregularities Q4									-0.100***	-0.085
									(0.035)	(0.058)
Facebook ads in low saturation						-0.004	0.007	0.008	0.008	0.031
						(0.018)	(0.022)	(0.021)	(0.022)	(0.023)
Facebook ads in low saturation × Above-median irregularities							-0.027	-0.031		
							(0.031)	(0.040)		
Facebook ads in low saturation × Irregularities Q3									-0.027	-0.089*
									(0.036)	(0.051)
Facebook ads in low saturation × Irregularities Q4									0.007	-0.018
									(0.036)	(0.056)
Spillover in high saturation						0.070***	0.125***	0.117***	0.119***	0.129***
						(0.022)	(0.022)	(0.025)	(0.022)	(0.026)
Spillover in high saturation × Above-median irregularities							-0.159***	-0.140***		
							(0.032)	(0.039)		
Spillover in high saturation × Irregularities Q3									-0.200***	-0.156***
									(0.036)	(0.043)
Spillover in high saturation × Irregularities Q4									-0.091***	-0.076
									(0.034)	(0.058)
Spillover in low saturation						-0.004	-0.003	0.003	-0.002	0.026
						(0.020)	(0.022)	(0.022)	(0.022)	(0.023)
Spillover in low saturation × Above-median irregularities							-0.000	-0.012		
							(0.033)	(0.041)		
Spillover in low saturation × Irregularities Q3									0.019	-0.051
									(0.037)	(0.051)
Spillover in low saturation × Irregularities Q4									0.024	-0.010
									(0.036)	(0.056)
Observations	13,254	13,254	13,254	13,254	13,254	13,254	13,254	13,254	13,254	13,254
R^2	0.61	0.53	0.72	0.66	0.72	0.66	0.71	0.78	0.73	0.78
Control outcome mean	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
Control outcome std. dev.	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
Test: null effect of Facebook ads below median (p value)		0.01	0.00	0.00	0.00					
Test: null effect of Facebook ads above median (p value)		0.29	0.04							
Test: null effect of Facebook ads in Q3 (p value)				0.06	0.07					
Test: null effect of Facebook ads in Q4 (p value)				0.55	0.67					
Test: null effect of spillover below median (p value)		0.23	0.06	0.10	0.03					
Test: null effect of spillover above median (p value)		0.86	0.04							
Test: null effect of spillover in Q3 (p value)				0.76	0.19					
Test: null effect of spillover in Q4 (p value)				0.79	0.36					
Test: same effect of Facebook ads in high and low (p value)						0.00				
Test: same effect of Facebook ads in high and low below median (p value)							0.00	0.00	0.00	0.00
Test: same effect of Facebook ads in high and low above median (p value)							0.00	0.03		
Test: same effect of Facebook ads in high and low in Q3 (p value)									0.05	0.37
Test: same effect of Facebook ads in high and low in Q4 (p value)									0.88	0.49
Test: same effect of spillovers in high and low (p value)						0.00				
Test: same effect of spillovers in high and low below median (p value)							0.00	0.00	0.00	0.00
Test: same effect of spillovers in high and low above median (p value)							0.00	0.01		
Test: same effect of spillovers in high and low in Q3 (p value)									0.02	0.70
Test: same effect of spillovers in high and low in Q4 (p value)									0.90	0.33
Treatment × covariate interactions			✓		✓			✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2. All precincts are weighted by the number of registered voters and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. The test statistics at the foot of the table report p values from two-sided hypothesis tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Table A11: Effect of Facebook ads on precinct-level municipal incumbent party vote share, weighting municipalities with different numbers of segments equally

	Incumbent party vote (share of turnout)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Facebook ads	0.020	0.055**	0.066**	0.051**	0.069**					
	(0.016)	(0.024)	(0.029)	(0.024)	(0.028)					
Facebook ads × Above-median irregularities		-0.075**	-0.059							
		(0.037)	(0.042)							
Facebook ads × Irregularities Q3				-0.081*	-0.037					
				(0.047)	(0.047)					
Facebook ads × Irregularities Q4				-0.051	-0.072					
				(0.040)	(0.045)					
Spillover	0.014	0.026	0.026	0.027	0.036					
	(0.017)	(0.023)	(0.026)	(0.022)	(0.025)					
Spillover × Above-median irregularities		-0.024	-0.017							
		(0.038)	(0.040)							
Spillover × Irregularities Q3				-0.002	0.008					
				(0.043)	(0.043)					
Spillover × Irregularities Q4				-0.039	-0.044					
				(0.047)	(0.046)					
Facebook ads in high saturation						0.020	0.068**	0.066*	0.063**	0.067*
						(0.019)	(0.029)	(0.037)	(0.028)	(0.036)
Facebook ads in high saturation × Above-median irregularities							-0.108**	-0.043		
							(0.043)	(0.057)		
Facebook ads in high saturation × Irregularities Q3									-0.122**	-0.045
									(0.056)	(0.059)
Facebook ads in high saturation × Irregularities Q4									-0.071	-0.027
									(0.046)	(0.065)
Facebook ads in low saturation						0.016	0.025	0.009	0.027	0.024
						(0.019)	(0.029)	(0.031)	(0.029)	(0.032)
Facebook ads in low saturation × Above-median irregularities							-0.018	-0.001		
							(0.044)	(0.046)		
Facebook ads in low saturation × Irregularities Q3									-0.001	-0.000
									(0.050)	(0.055)
Facebook ads in low saturation × Irregularities Q4									-0.031	-0.024
									(0.054)	(0.056)
Spillover in high saturation						0.017	0.065**	0.058	0.060**	0.058
						(0.019)	(0.029)	(0.039)	(0.029)	(0.037)
Spillover in high saturation × Above-median irregularities							-0.110**	-0.036		
							(0.044)	(0.059)		
Spillover in high saturation × Irregularities Q3									-0.126**	-0.042
									(0.059)	(0.061)
Spillover in high saturation × Irregularities Q4									-0.071	-0.014
									(0.045)	(0.067)
Spillover in low saturation						0.013	0.010	-0.006	0.012	0.009
						(0.020)	(0.028)	(0.031)	(0.027)	(0.031)
Spillover in low saturation × Above-median irregularities							0.008	0.032		
							(0.045)	(0.048)		
Spillover in low saturation × Irregularities Q3									0.046	0.052
									(0.050)	(0.056)
Spillover in low saturation × Irregularities Q4									-0.023	-0.008
									(0.054)	(0.057)
Observations	13,254	13,254	13,254	13,254	13,254	13,254	13,254	13,254	13,254	13,254
R^2	0.50	0.51	0.59	0.54	0.62	0.52	0.55	0.65	0.57	0.67
Control outcome mean	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
Control outcome std. dev.	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Test: null effect of Facebook ads below median (*p* value)		0.03	0.03	0.03	0.01					
Test: null effect of Facebook ads above median (*p* value)		0.39	0.76							
Test: null effect of Facebook ads in Q3 (*p* value)				0.43	0.35					
Test: null effect of Facebook ads in Q4 (*p* value)				0.99	0.93					
Test: null effect of spillover below median (*p* value)		0.26	0.32	0.22	0.15					
Test: null effect of spillover above median (*p* value)		0.94	0.70							
Test: null effect of spillover in Q3 (*p* value)				0.46	0.17					
Test: null effect of spillover in Q4 (*p* value)				0.78	0.82					
Test: same effect of Facebook ads in high and low (*p* value)						0.81				
Test: same effect of Facebook ads in high and low below median (*p* value)							0.13	0.07	0.20	0.12
Test: same effect of Facebook ads in high and low above median (*p* value)							0.03	0.40		
Test: same effect of Facebook ads in high and low in Q3 (*p* value)									0.05	0.97
Test: same effect of Facebook ads in high and low in Q4 (*p* value)									0.91	0.44
Test: same effect of spillovers in high and low (*p* value)						0.86				
Test: same effect of spillovers in high and low below median (*p* value)							0.04	0.05	0.08	0.09
Test: same effect of spillovers in high and low above median (*p* value)							0.00	0.19		
Test: same effect of spillovers in high and low in Q3 (*p* value)									0.03	0.56
Test: same effect of spillovers in high and low in Q4 (*p* value)									0.80	0.38
Treatment × covariate interactions			✓		✓			✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2. All municipalities are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. The test statistics at the foot of the table report *p* values from two-sided hypothesis tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Figure A5: Example of an infographic from a municipality (Hermosillo, Sonora) where irregularities were above 0%

A.8 Limited effects of the common knowledge variant of the Facebook ads

As noted in the main text, half the directly treated segments received Facebook ads with additional information designed to facilitate common knowledge about the campaign's scale. Figure A5 provides an example of this in the case of the 20% saturation campaign in the municipality of Hermosillo, Sonora. This slide was the penultimate slide in the video, and thus appeared right before the concluding slide (which contained no text).

As noted in the main text, the common knowledge treatment variant was pooled with the basic Facebook ads because we observed indistinguishable levels of engagement with these ads and no differential effects on voting behavior. First, Table A12 shows that the common knowledge treatment was no more likely to be engaged with than the non-common knowledge variant of the treatment. We also observe no statistically significant difference in user interactions for the common knowledge campaigns. Second, Table A13 further reports no notable difference in the effect of the common knowledge ("Facebook + CK") and non-common knowledge ("Facebook") variants of the Facebook ads. These findings may not be especially surprising, given that the common knowledge information only appeared toward the end of the ad or otherwise required that users read the comments or page associated with the ad.

A.9 Null effects of municipal level treatments on incumbent election victory

Table A14 reports no discernible effect of the municipal-level treatments on whether the municipal incumbent party was ultimately re-elected.

A.10 Results for other voting outcome measures

Table A15 reports the average and conditional treatment effects on incumbent vote share as a share of registered voters (rather than voters that turned out) for our main specifications. By using a pre-

Table A12: Effect of municipal treatments on municipal Facebook engagement, by common knowledge treatment

	Municipal counts per capita (normalized by 2015 adult population)								
	Paid-for impressions	Organic impressions	Paid-for unique viewers	Organic unique viewers	Unique user page engagements	Total views (of 3 seconds)	Unique views (of 3 seconds)	Total views (of entire video)	Unique views (of entire video)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Facebook ads (no common knowledge)	0.748***	0.038***	0.215***	0.021***	0.011***	0.143***	0.088***	0.035***	0.031***
	(0.092)	(0.006)	(0.025)	(0.004)	(0.001)	(0.017)	(0.010)	(0.004)	(0.003)
Facebook ads (common knowledge)	0.682***	0.036***	0.200***	0.020***	0.010***	0.132***	0.082***	0.032***	0.028***
	(0.094)	(0.007)	(0.025)	(0.004)	(0.002)	(0.017)	(0.010)	(0.004)	(0.004)
Observations	173	173	173	173	173	173	173	173	173
R^2	0.50	0.56	0.51	0.56	0.50	0.52	0.52	0.53	0.52
Control outcome mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Control outcome std. dev.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Test: No common knowledge = Common knowledge (p value, 2-sided)	0.511	0.747	0.567	0.669	0.426	0.562	0.558	0.416	0.429

Notes: Each specification is estimated using OLS, and includes randomization block fixed effects. Standard errors are clustered by municipality. The test statistics at the foot of the table report p values from two-sided hypothesis tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Table A13: Differential effects of Facebook ads on precinct-level municipal incumbent party vote share, by common knowledge treatment

	Incumbent party vote (share of turnout)					
	(1)	(2)	(3)	(4)	(5)	(6)
Facebook	0.018	0.050**	0.049**			
	(0.017)	(0.025)	(0.024)			
Facebook × Above-median irregularities		-0.078**				
		(0.036)				
Facebook × Irregularities Q3			-0.084*			
			(0.043)			
Facebook × Irregularities Q4			-0.047			
			(0.042)			
Facebook + CK	0.025	0.059**	0.056**			
	(0.017)	(0.025)	(0.024)			
Facebook + CK × Above-median irregularities		-0.076*				
		(0.040)				
Facebook + CK × Irregularities Q3			-0.058			
			(0.044)			
Facebook + CK × Irregularities Q4			-0.080*			
			(0.048)			
Spillover	0.019					
	(0.016)					
Spillover		0.048*	0.046*			
		(0.024)	(0.024)			
Spillover × Above-median irregularities		-0.064*				
		(0.038)				
Spillover × Irregularities Q3			-0.044			
			(0.044)			
Spillover × Irregularities Q4			-0.058			
			(0.043)			
Facebook ads in high saturation				0.031	0.078***	0.073***
				(0.019)	(0.028)	(0.027)
Facebook ads in high saturation × Above-median irregularities					-0.117***	
					(0.041)	
Facebook ads in high saturation × Irregularities Q3						-0.131**
						(0.052)
Facebook ads in high saturation × Irregularities Q4						-0.077*
						(0.045)
Facebook ads + CK in high saturation				0.031*	0.080***	0.075***
				(0.019)	(0.028)	(0.026)
Facebook ads in low saturation				0.004	0.019	0.023
				(0.021)	(0.030)	(0.030)
Facebook ads in low saturation × Above-median irregularities					-0.029	
					(0.041)	
Facebook ads in low saturation × Irregularities Q3						-0.034
						(0.045)
Facebook ads in low saturation × Irregularities Q4						-0.019
						(0.048)
Facebook ads + CK in low saturation				0.015	0.025	0.026
				(0.021)	(0.027)	(0.026)
Facebook ads + CK in low saturation × Above-median irregularities					-0.022	
					(0.045)	
Facebook ads + CK in low saturation × Irregularities Q3						0.011
						(0.046)
Facebook ads + CK in low saturation × Irregularities Q4						-0.065
						(0.060)
Spillover in high saturation				0.029	0.078***	0.073***
				(0.019)	(0.028)	(0.027)
Spillover in low saturation				0.007	0.009	0.011
				(0.019)	(0.026)	(0.026)
Spillover in low saturation × Above-median irregularities					0.002	
					(0.041)	
Spillover in low saturation × Irregularities Q3						0.050
						(0.045)
Spillover in low saturation × Irregularities Q4						-0.029
						(0.050)
Spillover in high saturation × Above-median irregularities					-0.124***	
					(0.041)	
Facebook ads + CK in high saturation × Above-median irregularities					-0.123***	
					(0.040)	
Spillover in high saturation × Irregularities Q3						-0.140**
						(0.054)
Spillover in high saturation × Irregularities Q4						-0.082*
						(0.043)
Facebook ads + CK in high saturation × Irregularities Q3						-0.137**
						(0.053)
Facebook ads + CK in high saturation × Irregularities Q4						-0.082*
						(0.043)
Observations	13,254	13,254	13,254	13,254	13,254	13,254
R^2	0.54	0.55	0.58	0.53	0.57	0.60
Control outcome mean	0.28	0.28	0.28	0.28	0.28	0.28
Control outcome std. dev.	0.14	0.14	0.14	0.14	0.14	0.14

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. The omitted irregularities category is Q1/2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. The test statistics at the foot of the table report p values from two-sided hypothesis tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Table A14: Effects of municipal saturation treatments on municipal election outcomes

	Incumbent party re-elected					
	(1)	(2)	(3)	(4)	(5)	(6)
Any Saturation	-0.081	-0.021	-0.030			
	(0.075)	(0.129)	(0.132)			
Any Saturation × Above-median irregularities		-0.128				
		(0.196)				
Any Saturation × Irregularities Q3			-0.059			
			(0.227)			
Any Saturation × Irregularities Q4			-0.140			
			(0.233)			
High saturation				-0.053	0.081	0.055
				(0.089)	(0.147)	(0.148)
High saturation × Above-median irregularities					-0.305	
					(0.219)	
High saturation × Irregularities Q3						-0.256
						(0.286)
High saturation × Irregularities Q4						-0.247
						(0.239)
Low saturation				-0.110	-0.138	-0.123
				(0.092)	(0.153)	(0.156)
Low saturation × Above-median irregularities					0.060	
					(0.231)	
Low saturation × Irregularities Q3						0.157
						(0.232)
Low saturation × Irregularities Q4						-0.033
						(0.285)
Observations	124	124	124	124	124	124
R^2	0.48	0.49	0.53	0.48	0.51	0.55
Control outcome mean	0.39	0.39	0.39	0.39	0.39	0.39
Control outcome std. dev.	0.49	0.49	0.49	0.49	0.49	0.49
Test: effect of treatment below median (p value, 2-sided)		0.87	0.82			
Test: effect of treatment above median (p value, 2-sided)		0.21				
Test: effect of treatment in Q3 (p value, 2-sided)			0.57			
Test: effect of treatment in Q4 (p value, 2-sided)			0.34			
Test: low = high (p value, 2-sided)				0.57		
Test: effect of low saturation below median (p value, 2-sided)					0.37	0.43
Test: effect of low saturation above median (p value, 2-sided)					0.59	
Test: effect of high saturation below median (p value, 2-sided)					0.58	0.71
Test: effect of high saturation above median (p value, 2-sided)					0.10	
Test: effect of low saturation in Q3 (p value, 2-sided)						0.81
Test: effect of low saturation in Q4 (p value, 2-sided)						0.48
Test: effect of high saturation in Q3 (p value, 2-sided)						0.36
Test: effect of high saturation in Q4 (p value, 2-sided)						0.28

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Robust standard errors are in parentheses. The test statistics at the foot of the table report p values from two-sided hypothesis tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

determined denominator, this outcome ensures that the estimates are not driven by potential effects of treatment on turnout. The point estimates are scaled down in accordance with the larger denominator, but the direction and statistically significance of the results are substantively unaffected.

Table A16 reports analogous results for turnout. The effects on turnout suggest that aggregate changes in turnout are not driving the aggregate changes in incumbent party vote share.

A.11 Testing for spillovers across municipalities

Table A17 reports the results of the tests for spillovers *across municipalities*, which are discussed in the main paper. The lack of large positive coefficients in Q1/2 provide little evidence to suggest that spillovers occurred across municipalities. However, these tests rely on a smaller sample of electoral precincts than our main analyses.

A.12 Removing randomization blocks

Figure A6 reports the results of our core regression specifications separately dropping each randomization block from the sample. For brevity, we report only the point estimates pertaining to our key findings documenting positive effects of the Facebook ads in directly and indirectly treated segments within municipalities with below-median levels of irregularities. The point estimates are robust to removing any particular randomization block.

A.13 Tests of Facebook mistargeting

As noted in the main text, we examine heterogeneity in treatment effects by Facebook's municipal-level movement range metric, which uses user location data to measure the average number of level-16 Bing tiles (which are approximately 600 meters by 600 meters in area at the equator) visited by Facebook users in a 24-hour period. This measure comes from Facebook's 2020 pre-pandemic Data For Good dataset. The results in Table A18 do not provide systematic evidence to suggest that the magnitude of indirect effects or the differential effects in high saturation municipalities increase with mobility, as inward mobility due to mistargeting would imply.

Our second mistargeting test leverages random variation in the location of segments that were part of the basic (NCK) and common knowledge (CK) ad campaigns within treated municipalities—specifically, high saturation municipalities as well as low saturation municipalities that contained 10 or more segments. Since treated segments were randomly assigned between these two versions of Borde Político's ad campaigns, the share of geographically adjacent treated segments that were part of a different campaign (e.g. where a segment receiving the CK campaign neighbors a segment receiving the NCK campaign) is also conditionally random. This allows us to estimate the effect of

Table A15: Effect of Facebook ads on precinct-level municipal incumbent party vote share, where the outcome denominator is the share of registered voters

	Incumbent party vote (share of registered voters)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Facebook ads	0.021*	0.044***	0.045**	0.040**	0.048**					
	(0.011)	(0.017)	(0.020)	(0.016)	(0.019)					
Facebook ads × Above-median irregularities		-0.056**	-0.042							
		(0.025)	(0.030)							
Facebook ads × Irregularities Q3				-0.045	-0.015					
				(0.032)	(0.032)					
Facebook ads × Irregularities Q4				-0.049*	-0.061*					
				(0.028)	(0.033)					
Spillover	0.009	0.022	0.019	0.023	0.026					
	(0.011)	(0.015)	(0.018)	(0.015)	(0.018)					
Spillover × Above-median irregularities		-0.028	-0.027							
		(0.025)	(0.028)							
Spillover × Irregularities Q3				-0.001	0.002					
				(0.026)	(0.028)					
Spillover × Irregularities Q4				-0.044	-0.053					
				(0.032)	(0.033)					
Facebook ads in high saturation						0.024*	0.054***	0.048*	0.049***	0.047*
						(0.013)	(0.019)	(0.025)	(0.018)	(0.024)
Facebook ads in high saturation × Above-median irregularities							-0.075***	-0.032		
							(0.029)	(0.041)		
Facebook ads in high saturation × Irregularities Q3									-0.069*	-0.026
									(0.037)	(0.041)
Facebook ads in high saturation × Irregularities Q4									-0.060*	-0.019
									(0.030)	(0.046)
Facebook ads in low saturation						0.007	0.021	0.004	0.023	0.015
						(0.013)	(0.019)	(0.022)	(0.018)	(0.022)
Facebook ads in low saturation × Above-median irregularities							-0.030	-0.025		
							(0.028)	(0.033)		
Facebook ads in low saturation × Irregularities Q3									-0.004	-0.007
									(0.029)	(0.036)
Facebook ads in low saturation × Irregularities Q4									-0.048	-0.055
									(0.035)	(0.039)
Spillover in high saturation						0.021	0.052***	0.045*	0.047**	0.044*
						(0.013)	(0.019)	(0.026)	(0.018)	(0.025)
Spillover in high saturation × Above-median irregularities							-0.078***	-0.034		
							(0.029)	(0.042)		
Spillover in high saturation × Irregularities Q3									-0.076*	-0.032
									(0.039)	(0.043)
Spillover in high saturation × Irregularities Q4									-0.059**	-0.014
									(0.030)	(0.047)
Spillover in low saturation						0.006	0.012	-0.003	0.014	0.008
						(0.013)	(0.018)	(0.022)	(0.017)	(0.022)
Spillover in low saturation × Above-median irregularities							-0.010	-0.003		
							(0.028)	(0.033)		
Spillover in low saturation × Irregularities Q3									0.027	0.023
									(0.029)	(0.036)
Spillover in low saturation × Irregularities Q4									-0.037	-0.040
									(0.036)	(0.040)
Observations	13,254	13,254	13,254	13,254	13,254	13,254	13,254	13,254	13,254	13,254
R^2	0.55	0.56	0.62	0.59	0.65	0.57	0.60	0.68	0.63	0.70
Control outcome mean	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
Control outcome std. dev.	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Test: null effect of Facebook ads below median (p value)		0.01	0.03	0.01	0.01					
Test: null effect of Facebook ads above median (p value)		0.46	0.90							
Test: null effect of Facebook ads in Q3 (p value)				0.84	0.19					
Test: null effect of Facebook ads in Q4 (p value)				0.69	0.54					
Test: null effect of spillover below median (p value)		0.15	0.30	0.12	0.14					
Test: null effect of spillover above median (p value)		0.75	0.63							
Test: null effect of spillover in Q3 (p value)				0.25	0.14					
Test: null effect of spillover in Q4 (p value)				0.45	0.26					
Test: same effect of Facebook ads in high and low (p value)						0.21				
Test: same effect of Facebook ads in high and low below median (p value)							0.06	0.03	0.13	0.07
Test: same effect of Facebook ads in high and low above median (p value)							0.08	0.85		
Test: same effect of Facebook ads in high and low in Q3 (p value)									0.19	0.71
Test: same effect of Facebook ads in high and low in Q4 (p value)									0.55	0.06
Test: same effect of spillovers in high and low (p value)						0.26				
Test: same effect of spillovers in high and low below median (p value)							0.01	0.02	0.03	0.05
Test: same effect of spillovers in high and low above median (p value)							0.01	0.43		
Test: same effect of spillovers in high and low in Q3 (p value)									0.11	0.81
Test: same effect of spillovers in high and low in Q4 (p value)									0.41	0.10
Treatment × covariate interactions			✓		✓			✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Specifications including interactive covariates further include interactions between each treatment condition and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. The test statistics at the foot of the table report p values from two-sided hypothesis tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Table A16: Effect of Facebook ads on precinct-level municipal turnout

	Turnout									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Facebook ads	0.013*	0.009	-0.011	0.008	-0.011					
	(0.008)	(0.012)	(0.009)	(0.012)	(0.009)					
Facebook ads × Above-median irregularities		0.005	0.036**							
		(0.018)	(0.015)							
Facebook ads × Irregularities Q3				0.024	0.049***					
				(0.020)	(0.015)					
Facebook ads × Irregularities Q4				-0.011	0.019					
				(0.022)	(0.019)					
Spillover	0.006	0.007	-0.011	0.006	-0.011					
	(0.008)	(0.013)	(0.009)	(0.013)	(0.009)					
Spillover × Above-median irregularities		-0.006	0.012							
		(0.017)	(0.014)							
Spillover × Irregularities Q3				0.005	0.024					
				(0.020)	(0.016)					
Spillover × Irregularities Q4				-0.017	-0.001					
				(0.020)	(0.017)					
Facebook ads in high saturation						0.013	0.006	-0.012	0.004	-0.014
						(0.009)	(0.013)	(0.010)	(0.012)	(0.010)
Facebook ads in high saturation × Above-median irregularities							0.013	0.050***		
							(0.020)	(0.016)		
Facebook ads in high saturation × Irregularities Q3									0.037	0.054***
									(0.023)	(0.017)
Facebook ads in high saturation × Irregularities Q4									-0.005	0.047**
									(0.025)	(0.021)
Facebook ads in low saturation						0.002	0.006	-0.010	0.006	-0.012
						(0.010)	(0.014)	(0.009)	(0.014)	(0.010)
Facebook ads in low saturation × Above-median irregularities							-0.011	0.011		
							(0.021)	(0.014)		
Facebook ads in low saturation × Irregularities Q3									0.003	0.036**
									(0.023)	(0.017)
Facebook ads in low saturation × Irregularities Q4									-0.025	-0.011
									(0.024)	(0.017)
Spillover in high saturation						0.008	0.002	-0.009	-0.000	-0.011
						(0.009)	(0.013)	(0.011)	(0.013)	(0.011)
Spillover in high saturation × Above-median irregularities							0.012	0.035**		
							(0.019)	(0.018)		
Spillover in high saturation × Irregularities Q3									0.032	0.037*
									(0.023)	(0.019)
Spillover in high saturation × Irregularities Q4									-0.001	0.034
									(0.023)	(0.023)
Spillover in low saturation						0.003	0.006	-0.006	0.007	-0.008
						(0.009)	(0.014)	(0.010)	(0.014)	(0.010)
Spillover in low saturation × Above-median irregularities							-0.011	0.001		
							(0.020)	(0.014)		
Spillover in low saturation × Irregularities Q3									-0.003	0.017
									(0.023)	(0.017)
Spillover in low saturation × Irregularities Q4									-0.020	-0.013
									(0.024)	(0.017)
Observations	13,254	13,254	13,254	13,254	13,254	13,254	13,254	13,254	13,254	13,254
R^2	0.63	0.63	0.69	0.63	0.69	0.64	0.65	0.71	0.65	0.71
Control outcome mean	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64
Control outcome std. dev.	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
Test: null effect of Facebook ads below median (p value)		0.46	0.18	0.52	0.20					
Test: null effect of Facebook ads above median (p value)		0.21	0.02							
Test: null effect of Facebook ads in Q3 (p value)				0.03	0.00					
Test: null effect of Facebook ads in Q4 (p value)				0.85	0.62					
Test: null effect of spillover below median (p value)		0.60	0.24	0.63	0.25					
Test: null effect of spillover above median (p value)		0.97	0.83							
Test: null effect of spillover in Q3 (p value)				0.43	0.23					
Test: null effect of spillover in Q4 (p value)				0.46	0.32					
Test: same effect of Facebook ads in high and low (p value)						0.18				
Test: same effect of Facebook ads in high and low below median (p value)							0.97	0.80	0.82	0.84
Test: same effect of Facebook ads in high and low above median (p value)							0.20	0.00		
Test: same effect of Facebook ads in high and low in Q3 (p value)									0.05	0.20
Test: same effect of Facebook ads in high and low in Q4 (p value)									0.47	0.00
Test: same effect of spillovers in high and low (p value)						0.44				
Test: same effect of spillovers in high and low below median (p value)							0.68	0.64	0.50	0.67
Test: same effect of spillovers in high and low above median (p value)							0.18	0.02		
Test: same effect of spillovers in high and low in Q3 (p value)									0.10	0.59
Test: same effect of spillovers in high and low in Q4 (p value)									0.46	0.09
Treatment × covariate interactions			✓		✓			✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

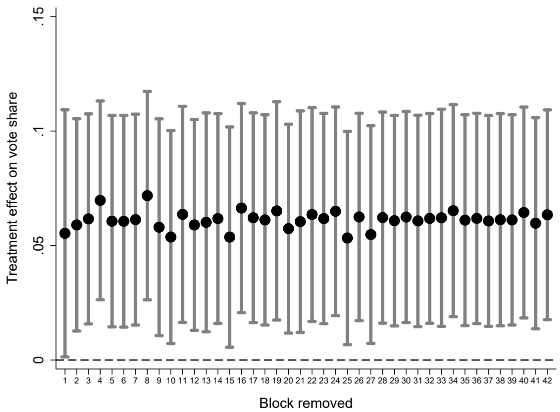
Table A17: Effect of Facebook ads on precinct-level vote share in nearby municipalities

	Vote for incumbent party in nearby experimental municipality (share of turnout)				
	(1)	(2)	(3)	(4)	(5)
Facebook ads	0.034	0.011	-0.013	0.012	-0.003
	(0.036)	(0.054)	(0.046)	(0.055)	(0.044)
Facebook ads × Above-median irregularities		0.007	0.069		
		(0.067)	(0.065)		
Facebook ads × Irregularities Q3				-0.044	0.013
				(0.076)	(0.071)
Facebook ads × Irregularities Q4				0.062	0.127*
				(0.076)	(0.070)
Spillover	0.034	-0.018	-0.013	-0.017	-0.003
	(0.040)	(0.060)	(0.052)	(0.060)	(0.051)
Spillover × Above-median irregularities		0.087	0.064		
		(0.076)	(0.067)		
Spillover × Irregularities Q3				0.056	0.030
				(0.088)	(0.077)
Spillover × Irregularities Q4				0.118	0.078
				(0.092)	(0.095)
Observations	640	640	640	640	640
R^2	0.09	0.12	0.39	0.13	0.40
Control outcome mean	0.24	0.24	0.24	0.24	0.24
Control outcome std. dev.	0.16	0.16	0.16	0.16	0.16
Test: null effect of Facebook ads below median (p value)		0.83	0.77	0.83	0.94
Test: null effect of Facebook ads above median (p value)		0.66	0.16		
Test: null effect of Facebook ads in Q3 (p value)				0.55	0.84
Test: null effect of Facebook ads in Q4 (p value)				0.17	0.04
Test: null effect of spillover below median (p value)		0.77	0.80	0.77	0.96
Test: null effect of spillover above median (p value)		0.16	0.20		
Test: null effect of spillover in Q3 (p value)				0.57	0.63
Test: null effect of spillover in Q4 (p value)				0.14	0.35
Treatment × covariate interactions			✓		✓

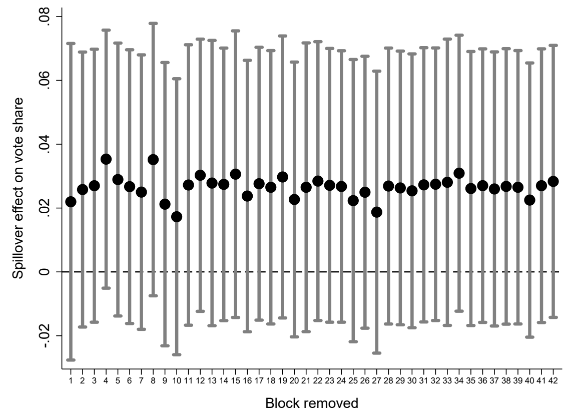
Notes: Each specification is estimated using OLS, and includes a lagged dependent variable. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2. All observations are weighted by the inverse probability of treatment assignment and weight each experimental segment equally. Standard errors clustered by municipality are in parentheses. The test statistics at the foot of the table report p values from two-sided hypothesis tests. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Figure A6: Leave-one-out estimates (dropping each randomization block) of the average treatment effect of Facebook ads on incumbent party vote share in the municipalities where below-median irregularities occurred

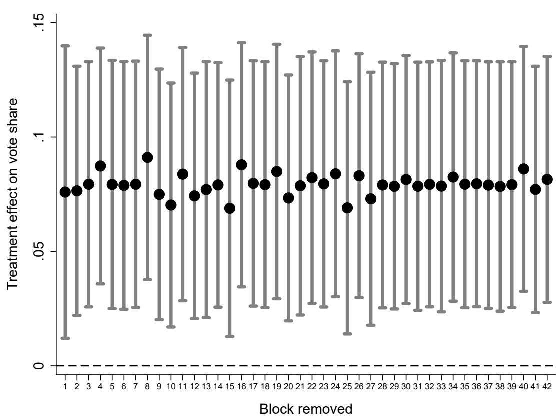
(a) Conditional average treatment effect



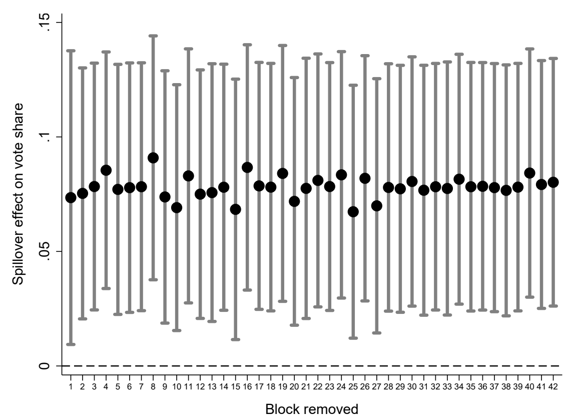
(b) Conditional average spillover effect



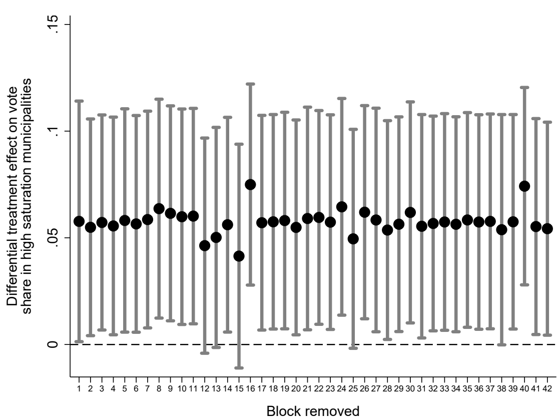
(c) Conditional average treatment effect in high saturation municipalities



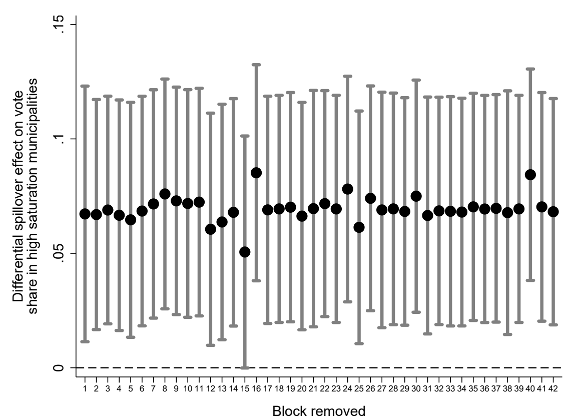
(d) Conditional average spillover effect in high saturation municipalities



(e) Difference in conditional average treatment effect between high and low saturation municipalities



(f) Difference in conditional spillover treatment effect between high and low saturation municipalities



Notes: The estimates in subfigures (a) and (b) derive from regressions analogous to column (2) of Table 2. The estimates in subfigures (c)-(f) derive from regressions analogous to column (2) of Table 3. Each estimate corresponds to dropping the block arrayed along the horizontal axis. Vertical bars denote 95% confidence intervals.

Table A18: Effect of Facebook ads spillovers on precinct-level municipal incumbent party vote share, by municipal mobility range

	Incumbent party vote (share of turnout)	
	(1)	(2)
Facebook ads in high saturation × Movement range	0.048	0.031
	(0.050)	(0.046)
Facebook ads in high saturation × Above-median irregularities × Movement range	-0.131**	
	(0.058)	
Facebook ads in high saturation × Irregularities Q3 × Movement range		-0.077
		(0.060)
Facebook ads in high saturation × Irregularities Q4 × Movement range		-0.029
		(0.089)
Facebook ads in low saturation × Movement range	0.008	0.000
	(0.032)	(0.029)
Facebook ads in low saturation × Above-median irregularities × Movement range	0.036	
	(0.046)	
Facebook ads in low saturation × Irregularities Q3 × Movement range		0.030
		(0.049)
Facebook ads in low saturation × Irregularities Q4 × Movement range		0.179**
		(0.069)
Spillover in high saturation × Movement range	0.044	0.027
	(0.050)	(0.046)
Spillover in high saturation × Above-median irregularities × Movement range	-0.113*	
	(0.058)	
Spillover in high saturation × Irregularities Q3 × Movement range		-0.049
		(0.060)
Spillover in high saturation × Irregularities Q4 × Movement range		-0.030
		(0.086)
Spillover in low saturation × Movement range	0.019	0.011
	(0.032)	(0.029)
Spillover in low saturation × Above-median irregularities × Movement range	0.031	
	(0.045)	
Spillover in low saturation × Irregularities Q3 × Movement range		0.027
		(0.049)
Spillover in low saturation × Irregularities Q4 × Movement range		0.099
		(0.089)
Observations	12,470	12,470
R^2	0.67	0.70
Control outcome mean	0.28	0.28
Control outcome std. dev.	0.14	0.14
Mean movement range	-0.04	-0.04

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Lower-order interaction terms are included, but omitted to save space. The omitted irregularities category is Q1/2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

a given treated segment being near to more treated segments that are part of a different campaign by estimating the following regression among directly treated segments:

$$\begin{aligned} Y_{psm} &= \alpha Y_{psm}^{lag} + \beta \textit{Share of adjacent treated segments in a different ad campaign}_{sm} \\ &\quad + \eta_n + \mu_m + \varepsilon_{psm}, \end{aligned} \tag{A1}$$

where η_n are fixed effects for the number of directly treated adjacent segments and μ_m are municipality fixed effects. Our main tests further examine heterogeneity by irregularities quartile by including interactions with *Share of adjacent treated segments in a different ad campaign*$_{sm}$ and η_n. Since greater exposure to ads due to potential mistargeting within the vicinity of treated segments should be greater where adjacent segments were part of a different campaign (as explained in the main paper), we would expect mistargeting to imply bigger treatment effects in the directly treated segments that were adjacent to more segments that were part of a different ad campaign, if mistargeting is driving the estimates. The results in Appendix Table A19 report no evidence to suggest that mistargeting of this form drives the results.

A.14 Changes in the effective number of parties

We use the ENPV to explore whether Facebook ads, especially in high saturation campaigns, coordinate voting behavior. A decrease in the ENPV would be consistent with votes concentrating among fewer parties, as theories of tacit and explicit coordination predict.

However, it is hard to distinguish vote coordination from sincere changes in support for parties that are induced by other mechanisms, such as belief updating. This is because any change in vote share will alter the ENPV. To sign the mechanical effect of changes in support for the incumbent party on the ENPV due to pure belief updating, we formally analyze the properties of the ENPV metric. Let $s_{jpm} \in [0,1]$ be the pre-intervention vote share of party $j = 1,...,J$ in precinct p of municipality m; vote shares across parties satisfy $\sum_j s_{jpm} = 1$. To allow for the treatment to affect party vote shares, denote the incumbent party as $j = 1$ and let $v_{1pm} = s_{1pm} + \Delta_{pm}$, where Δ_{pm} captures changes in the incumbent party's vote share due to treatment. Since we are using vote shares, changes in support for the incumbent party entail changes in support for other parties as well; specifically, the vote share of any party $j \neq 1$ is given by $v_{jpm} = s_{jpm} - \alpha_{jpm}\Delta_{pm}$, where $\alpha_{jpm} \in [0,1]$ captures the degree to which votes are lost/gained by party j and $\sum_{j\neq 1} \alpha_{jpm} = 1$.

Table A19: Effect of adjacent directly treated segments being part of a different campaign on precinct-level municipal incumbent party vote share

	Incumbent party vote (share of turnout)	
	(1)	(2)
Share of adjacent treated segments in a different ad campaign	-0.007	-0.007
	(0.007)	(0.007)
Share of adjacent treated segments in a different ad campaign × Above median irregularities	-0.003	
	(0.012)	
Share of adjacent treated segments in a different ad campaign × Irregularities Q3		-0.020
		(0.015)
Share of adjacent treated segments in a different ad campaign × Irregularities Q4		0.023**
		(0.010)
Observations	1,722	1,722
R^2	0.75	0.75
Outcome mean	0.65	0.65
Outcome std. dev.	0.10	0.10
Share of adjacent treated segments in a different ad campaign mean	0.662	0.662
Share of adjacent treated segments in a different ad campaign std. dev.	0.418	0.418

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable, fixed effects for the number of nearby segments that are treated, and municipality fixed effects. The sample is restricted to precincts in high saturation municipalities and precincts in low saturation municipalities that contained 10 or more segments. The analysis focuses on precincts (in treated segments) within 100 meters of another treated segment. The omitted irregularities category is Q1/2. All segments are weighted equally. Standard errors clustered by municipality are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Differentiating the ENPV with respect to Δ_{pm} then yields the following result:

$$\begin{aligned}\frac{\partial ENPV_{pm}}{\partial \Delta_{pm}} &= \frac{\partial}{\partial \Delta_{pm}} \left(\sum_{j=1}^{J} v_{jpm}^2 \right)^{-1} \\ &= -\frac{2\left[s_{1pm} - \sum_{j\neq 1} \alpha_{jpm} s_{jpm} + \Delta_{pm}\left(1 + \sum_{j\neq 1} \alpha_{jpm}^2\right)\right]}{\left(\sum_{j=1}^{J} v_{jpm}^2\right)^2} \\ &= \begin{cases} > 0 & \text{if } s_{1pm} < \bar{s}_{1pm} := \sum_{j\neq 1} \alpha_{jpm} s_{jpm} - \Delta_{pm}\left(1 + \sum_{j\neq 1} \alpha_{jpm}^2\right) \\ \leq 0 & \text{otherwise} \end{cases} \end{aligned} \tag{A2}$$

which implies that increases in incumbent party vote share induced by treatment increase the ENPV when the incumbent party's baseline level of support, s_{1pm}, is sufficiently low relative to the parties that they gain vote share from. In our empirical application, the average incumbent party received 28% of votes in our control group. This is likely to fall below $\bar{s}_{1pm}$, as the median incumbent, in the same group of municipalities, won her race with 43% of votes in 2015. This is likely because incumbent parties in Q1/2 mostly gain votes from the other main parties, suggesting a high α_{jpm} for one or two other popular parties.[6]

Although the formal analysis was applied to precinct-level voting decisions, coordination could occur—especially if is driven by common knowledge—at higher levels as well. To identify the effect of segment-level treatments, we examine effects on ENPV defined at the precinct and segment levels. The results in Tables A20 and A21, for precinct- and segment-level ENPV measures respectively, provide evidence consistent with voter coordination. Direct and indirect exposure to Facebook ads, primarily in high saturation municipalities, significantly reduced the ENPV in Q1/2. The effect becomes positive or null in Q3 and Q4. Because of our weighting and clustering, the segment-level ENPV analysis is analogous to collapsing the precinct-level data to the segment level.

A.15 Limited online responses to the Facebook ads

We collected all comments on, reactions to, and shares of Borde Político's Facebook ads or any other content posted on Borde Político's Facebook pages. For each of these 9,460 interactions, we collected the username, profile URL, type of interaction (comment, reaction, and/or share), and content of the comment (if any). Interactions consist of 1,390 comments, 6,567 reactions (4,261 "Like", 1,092 "Angry", 569 "Wow", 515 "Haha", 85 "Love", and 45 "Sad"), and 1,503 shares.

[6] Strategic voters generally intend to vote for front runners or runners up, and those are the ones likely to change vote intention towards the incumbent as a result of coordination.

Table A20: Effect of Facebook ads on precinct-level ENPV, by information campaign saturation

	ENPV (precinct)				
	(1)	(2)	(3)	(4)	(5)
Facebook ads in high saturation	-0.110	-0.345*	-0.423**	-0.322	-0.310*
	(0.123)	(0.204)	(0.194)	(0.195)	(0.182)
Facebook ads in high saturation × Above-median irregularities		0.677**	0.930***		
		(0.288)	(0.315)		
Facebook ads in high saturation × Irregularities Q3				0.890**	1.074***
				(0.366)	(0.307)
Facebook ads in high saturation × Irregularities Q4				0.407	0.459
				(0.305)	(0.360)
Facebook ads in low saturation	-0.198	-0.084	0.072	-0.083	0.109
	(0.126)	(0.200)	(0.174)	(0.193)	(0.167)
Facebook ads in low saturation × Above-median irregularities		-0.235	-0.321		
		(0.274)	(0.258)		
Facebook ads in low saturation × Irregularities Q3				-0.684**	-0.980***
				(0.300)	(0.269)
Facebook ads in low saturation × Irregularities Q4				0.104	0.209
				(0.333)	(0.316)
Spillover in high saturation	-0.132	-0.318	-0.407**	-0.294	-0.301*
	(0.123)	(0.207)	(0.188)	(0.198)	(0.175)
Spillover in high saturation × Above-median irregularities		0.535*	0.800**		
		(0.286)	(0.307)		
Spillover in high saturation × Irregularities Q3				0.693*	0.876***
				(0.379)	(0.314)
Spillover in high saturation × Irregularities Q4				0.312	0.425
				(0.300)	(0.349)
Spillover in low saturation	-0.207*	-0.068	0.049	-0.066	0.091
	(0.123)	(0.196)	(0.181)	(0.188)	(0.174)
Spillover in low saturation × Above-median irregularities		-0.286	-0.335		
		(0.264)	(0.255)		
Spillover in low saturation × Irregularities Q3				-0.725**	-0.925***
				(0.314)	(0.278)
Spillover in low saturation × Irregularities Q4				0.043	0.132
				(0.311)	(0.305)
Observations	13,254	13,254	13,254	13,254	13,254
R^2	0.41	0.46	0.59	0.49	0.62
Control outcome mean	3.39	3.39	3.39	3.39	3.39
Control outcome std. dev.	0.83	0.83	0.83	0.83	0.83
Test: same effect of Facebook ads in high and low (p value)	0.44				
Test: same effect of Facebook ads in high and low below median (p value)		0.07	0.00	0.10	0.01
Test: same effect of Facebook ads in high and low above median (p value)		0.00	0.00		
Test: same effect of Facebook ads in high and low in Q3 (p value)				0.00	0.00
Test: same effect of Facebook ads in high and low in Q4 (p value)				0.79	0.56
Test: same effect of spillovers in high and low (p value)	0.50				
Test: same effect of spillovers in high and low below median (p value)		0.08	0.00	0.11	0.01
Test: same effect of spillovers in high and low above median (p value)		0.00	0.00		
Test: same effect of spillovers in high and low in Q3 (p value)				0.00	0.00
Test: same effect of spillovers in high and low in Q4 (p value)				0.93	0.98
Treatment × covariate interactions			✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Lower-order interaction terms are included, but omitted to save space. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2 in columns (2)-(5). All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Table A21: Effect of Facebook ads on segment-level ENPV, by information campaign saturation

	ENPV (segment)				
	(1)	(2)	(3)	(4)	(5)
Facebook ads in high saturation	-0.144	-0.383*	-0.457**	-0.359*	-0.330*
	(0.135)	(0.222)	(0.208)	(0.214)	(0.195)
Facebook ads in high saturation × Above-median irregularities		0.702**	0.945***		
		(0.315)	(0.342)		
Facebook ads in high saturation × Irregularities Q3				0.962**	1.108***
				(0.403)	(0.338)
Facebook ads in high saturation × Irregularities Q4				0.386	0.411
				(0.330)	(0.397)
Facebook ads in low saturation	-0.241*	-0.077	0.080	-0.076	0.110
	(0.139)	(0.218)	(0.195)	(0.211)	(0.187)
Facebook ads in low saturation × Above-median irregularities		-0.339	-0.411		
		(0.297)	(0.286)		
Facebook ads in low saturation × Irregularities Q3				-0.831**	-1.127***
				(0.335)	(0.304)
Facebook ads in low saturation × Irregularities Q4				0.019	0.143
				(0.354)	(0.347)
Spillover in high saturation	-0.146	-0.365	-0.428**	-0.341	-0.307
	(0.138)	(0.225)	(0.203)	(0.216)	(0.190)
Spillover in high saturation × Above-median irregularities		0.639*	0.872**		
		(0.326)	(0.337)		
Spillover in high saturation × Irregularities Q3				0.869*	0.975***
				(0.461)	(0.354)
Spillover in high saturation × Irregularities Q4				0.343	0.412
				(0.318)	(0.382)
Spillover in low saturation	-0.245*	-0.059	0.055	-0.055	0.100
	(0.135)	(0.214)	(0.199)	(0.207)	(0.193)
Spillover in low saturation × Above-median irregularities		-0.385	-0.422		
		(0.288)	(0.282)		
Spillover in low saturation × Irregularities Q3				-0.842**	-1.020***
				(0.351)	(0.315)
Spillover in low saturation × Irregularities Q4				-0.055	0.020
				(0.331)	(0.338)
Observations	13,254	13,254	13,254	13,254	13,254
R^2	0.47	0.52	0.68	0.57	0.71
Control outcome mean	3.51	3.51	3.51	3.51	3.51
Control outcome std. dev.	0.81	0.81	0.81	0.81	0.81
Test: same effect of Facebook ads in high and low (p value)	0.44				
Test: same effect of Facebook ads in high and low below median (p value)		0.05	0.00	0.06	0.01
Test: same effect of Facebook ads in high and low above median (p value)		0.00	0.00		
Test: same effect of Facebook ads in high and low in Q3 (p value)				0.00	0.00
Test: same effect of Facebook ads in high and low in Q4 (p value)				0.74	0.59
Test: same effect of spillovers in high and low (p value)	0.42				
Test: same effect of spillovers in high and low below median (p value)		0.05	0.00	0.06	0.01
Test: same effect of spillovers in high and low above median (p value)		0.00	0.00		
Test: same effect of spillovers in high and low in Q3 (p value)				0.00	0.00
Test: same effect of spillovers in high and low in Q4 (p value)				0.85	0.81
Treatment × covariate interactions			✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Lower-order interaction terms are included, but omitted to save space. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2 in columns (2)-(5). All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

When a Facebook profile was public, three coders further collected information about partisan affiliation and whether users were operatives of a political campaign, worked at a political party, or were candidates themselves. Additionally, we cross-checked the list of Facebook users using the universe of candidates and substitutes running in the 2018 election across all election types (i.e. city councilor, mayor, state deputy, federal deputy, senator, governor, and president), which accounts for almost 17,000 individuals, to identify candidates reacting to Borde Político's content.

In total, we found four candidates reacting to Borde Político's ads:

1. A PAN candidate for federal deputy who liked an ad reporting 30% irregularities by the PRI government in Ciudad Valles, San Luis Potosí;

2. A Citizen's Movement (MC) candidate for municipal president who shared an ad reporting 14% irregularities by the PRD government in Venustiano Carranza, Michoacán (the candidate was eventually elected);

3. An Alternative Sonora Movement candidate for municipal president who shared an ad reporting 30% irregularities by the PRI government in Huatabampo, Sonora; and

4. A PRD candidate for federal deputy who commented on an ad reporting 61% irregularities in the PRD municipal government in Cuautla, Morelos (the candidate challenged the information and argued that there were no irregularities).

Furthermore, we did not find systematic or coordinated reactions from political party operatives or partisans either: we only found 8 reactions coming from potential party operatives who worked at a political campaign or political party.

We also looked for online responses to Borde Político's ads on Twitter. Using the list of names of the 892 candidates running for mayor in municipalities in our sample, we identified candidates' personal and professional Twitter accounts. We detected a total of 278 Twitter accounts belonging to a candidate or mayoral campaign.[7] We then scraped all tweets that were posted by each one of these accounts and looked for 18 keywords related to Borde Político's ad campaign: *ASF, Auditoría Superior de la Federación, FISM, Fondo de Aportaciones para la Infraestructura Social Municipal, FAIS, Fondo de Aportaciones para la Infraestructura Social, Borde Político, Facebook, Corrupción, Inobservancias en el ejercicio de los recursos municipales, Inobservancias + recursos municipales, ZAP, Zonas de Atención Prioritaria, No Benefician, Población en pobreza extrema, Población en rezago social, No Autorizados, Observar la normativa*. Additionally, we ran an open

[7] While Facebook has high penetration in Mexico, Twitter has been lagging behind. According to ENDUTIH, in 2018, only 9.7 million adults (around 12%) used Twitter at least once a week compared to 56.1 millions (around 70%) who used Facebook.

search of these keywords on Twitter to see if there were reactions or content sharing more generally by unidentified candidates or any other user. We did not find any direct or indirect reaction to Borde's Político ad campaign on Twitter.

A.16 No evidence of media reporting driving the effects

We collected online data from 263 local newspapers serving 92 of the 124 municipalities in our final sample. First, we compiled a list of all local newspapers serving either control or treated municipalities in our sample using the national register of newspapers ("Padrón Nacional de Medios Impresos;" https://pnmi.segob.gob.mx). While the original list consisted of more than 800 newspapers, we filtered news outlets according to three criteria: (i) serving a municipality in our sample; (ii) were open-access (dropping 6 small newspapers with a paywall); and (iii) having a working webpage where original content was posted. Second, we scraped all 263 websites separately and obtained all articles published between February and July, 2018 (election day was on July 1st, 2018). Third, we classified all articles containing at least one of the 18 terms related to Borde Político's ad campaign as we did with Twitter accounts belonging to mayoral candidates (see Section A.15). Fourth, with the help of three manual coders we read and classified each article into four categories: direct mentioning of Borde Político's ad campaign, indirect mentioning of Borde Político's ad campaign, mentions of corruption in general, and none of the above.

We only found one case directly mentioning Borde Político (see here), but the article, published 11 days after the election, describes a separate foregoing campaign by Borde Político designed to hold federal deputies and senators accountable. We also identified 6 mentions of either the ASF or local governments' spending irregularities, but none of them related to Borde Político's ad campaign. Only one of these articles was published after Borde Político's ad campaign was started and it was published two days after the ad campaign ended, the day before the election. Finally, we found 33 mentions of corruption in general; 29 of these articles were published while Borde Político's ad campaign was running, but none of them related directly or indirectly to the information disseminated by the Facebook ads.

More formally, Table A22 reports our estimates of the effect of municipal-level saturation interventions on local newspaper coverage of corruption in general. The results, estimated using equation (4) in the main paper, show no evidence to suggest that reporting on corruption increased in the days after the launch of Facebook ads but before election day. This holds when considering both the total number of articles written on corruption, as well as the circulation-adjusted total number of articles.

The preceding analyses relate to newspapers, rather than broadcast media outlets. Since transcripts of radio and local television programming were not available, we consider access to local radio and television stations—that could have reported on municipal malfeasance, either after or

Table A22: Effect of Facebook ad campaigns on municipal-level newspaper reporting on corruption in general

	Total articles on corruption										Total articles on corruption (circulation-weighted)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Any Saturation	-6.209	-5.780	-13.285	-5.483	-14.962						-6.209	-5.826	-13.372	-5.533	-15.042					
	(4.393)	(8.355)	(11.730)	(8.429)	(12.038)						(4.382)	(8.353)	(11.717)	(8.429)	(12.028)					
Any Saturation × Above-median irregularities		-2.635	9.577									-2.544	9.701							
		(10.517)	(17.338)									(10.485)	(17.313)							
Any Saturation × Irregularities Q3				-10.003	2.179									-9.803	2.352					
				(13.033)	(15.896)									(12.964)	(15.859)					
Any Saturation × Irregularities Q4				4.608	25.325									4.593	25.391					
				(9.819)	(22.613)									(9.808)	(22.606)					
High saturation						-4.125	-4.419	-17.898	-4.548	-24.509						-4.173	-4.469	-17.927	-4.609	-24.494
						(4.746)	(9.895)	(15.549)	(9.965)	(17.266)						(4.735)	(9.895)	(15.535)	(9.965)	(17.256)
High saturation × Above-median irregularities							-1.539	23.856									-1.551	23.849		
							(12.767)	(24.615)									(12.735)	(24.588)		
High saturation × Irregularities Q3									-3.511	17.325									-3.390	17.381
									(13.155)	(23.292)									(13.121)	(23.260)
High saturation × Irregularities Q4									0.998	54.470									0.875	54.246
									(12.537)	(33.992)									(12.504)	(33.982)
Low saturation						-8.370*	-7.385	-15.450	-6.636	-18.081						-8.321*	-7.428	-15.521	-6.676	-18.112
						(4.687)	(7.138)	(14.527)	(7.282)	(14.416)						(4.675)	(7.139)	(14.508)	(7.283)	(14.409)
Low saturation × Above-median irregularities							-3.068	12.050									-2.886	12.237		
							(9.601)	(20.443)									(9.564)	(20.410)		
Low saturation × Irregularities Q3									-15.577	3.257									-15.305	3.449
									(15.274)	(19.237)									(15.174)	(19.190)
Low saturation × Irregularities Q4									7.718	35.836									7.790	35.886
									(8.791)	(27.853)									(8.800)	(27.856)
Observations	124	124	124	124	124	124	124	124	124	124	124	124	124	124	124	124	124	124	124	124
R^2	0.37	0.38	0.58	0.40	0.60	0.37	0.39	0.64	0.42	0.68	0.37	0.38	0.58	0.40	0.60	0.37	0.39	0.64	0.42	0.68
Control outcome mean	8.51	8.51	8.51	8.51	8.51	8.51	8.51	8.51	8.51	8.51	8.46	8.46	8.46	8.46	8.46	8.46	8.46	8.46	8.46	8.46
Control outcome std. dev.	31.53	31.53	31.53	31.53	31.53	31.53	31.53	31.53	31.53	31.53	31.45	31.45	31.45	31.45	31.45	31.45	31.45	31.45	31.45	31.45
Test: effect of treatment below median (p value)		0.49	0.26	0.52	0.22							0.49	0.26	0.51	0.22					
Test: effect of treatment above median (p value)		0.16	0.63									0.16	0.63							
Test: effect of treatment in Q3 (p value)				0.13	0.16									0.13	0.17					
Test: effect of treatment in Q4 (p value)				0.79	0.40									0.78	0.40					
Test: low = high (p value)						0.21										0.22				
Test: effect of low saturation below median (p value)							0.30	0.29	0.37	0.22							0.30	0.29	0.36	0.22
Test: effect of low saturation above median (p value)							0.29	0.58									0.28	0.58		
Test: effect of low saturation in Q3 (p value)									0.12	0.22									0.13	0.23
Test: effect of low saturation in Q4 (p value)									0.71	0.25									0.70	0.25
Test: effect of high saturation in Q3 (p value)							0.66	0.26	0.65	0.16							0.65	0.25	0.65	0.16
Test: effect of high saturation in Q4 (p value)							0.34	0.61									0.33	0.61		
F_test_Q3_2									0.22	0.59									0.21	0.59
F_test_Q4_2									0.61	0.11									0.59	0.11
Treatment × covariate interactions			✓		✓			✓		✓			✓		✓			✓		✓

Notes: Each specification is estimated using OLS, and includes randomization block fixed effects. Lower-order interaction terms are included, but omitted to save space. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2. The omitted irregularities category is Q1/2. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

before Borde Político's campaign—as potential moderators. Tables A23 and A24 examine heterogeneity in the effects of Facebook ads by the number of local and non-local media outlets covering a given precinct. However, the results provide no evidence to suggest that the magnitude of treatment effects is increasing in media coverage. This suggests that media coverage did not complement the dissemination of Facebook ads. If anything, media serves as a substitute, although the estimates are not systematically statistically significant.

Table A23: Differential effect of Facebook ads on precinct-level municipal incumbent party vote share, by the number of local media outlets at precinct is covered by

	Incumbent party vote (share of turnout)			
	(1)	(2)	(3)	(4)
Facebook ads in high saturation × Local media outlets	-0.002	-0.010*	-0.001	-0.002
	(0.003)	(0.006)	(0.003)	(0.006)
Facebook ads in high saturation × Above-median irregularities × Local media outlets	0.001	-0.001		
	(0.005)	(0.006)		
Facebook ads in high saturation × Irregularities Q3 × Local media outlets			-0.000	-0.009
			(0.007)	(0.008)
Facebook ads in high saturation × Irregularities Q4 × Local media outlets			0.006	0.016**
			(0.008)	(0.007)
Facebook ads in low saturation × Local media outlets	-0.004*	0.000	-0.004*	0.004
	(0.002)	(0.006)	(0.002)	(0.006)
Facebook ads in low saturation × Above-median irregularities × Local media outlets	0.003	-0.006		
	(0.004)	(0.005)		
Facebook ads in low saturation × Irregularities Q3 × Local media outlets			0.001	-0.017**
			(0.004)	(0.008)
Facebook ads in low saturation × Irregularities Q4 × Local media outlets			0.009*	0.003
			(0.006)	(0.006)
Spillover in high saturation × Local media outlets	-0.002	-0.010*	-0.001	-0.002
	(0.003)	(0.006)	(0.003)	(0.006)
Spillover in high saturation × Above-median irregularities × Local media outlets	0.002	-0.003		
	(0.005)	(0.006)		
Spillover in high saturation × Irregularities Q3 × Local media outlets			-0.002	-0.013*
			(0.007)	(0.007)
Spillover in high saturation × Irregularities Q4 × Local media outlets			0.010	0.015**
			(0.007)	(0.007)
Spillover in low saturation × Local media outlets	-0.004	0.001	-0.003	0.005
	(0.002)	(0.005)	(0.002)	(0.006)
Spillover in low saturation × Above-median irregularities × Local media outlets	0.002	-0.007		
	(0.004)	(0.005)		
Spillover in low saturation × Irregularities Q3 × Local media outlets			0.001	-0.018**
			(0.004)	(0.008)
Spillover in low saturation × Irregularities Q4 × Local media outlets			0.009	0.003
			(0.005)	(0.005)
Observations	13,234	13,234	13,234	13,234
R^2	0.59	0.69	0.61	0.71
Control outcome mean	0.28	0.28	0.28	0.28
Control outcome std. dev.	0.14	0.14	0.14	0.14
Treatment × covariate interactions		✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Lower-order interaction terms are included, but omitted to save space. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).

Table A24: Differential effect of Facebook ads on precinct-level municipal incumbent party vote share, by the number of non-local media outlets at precinct is covered by

	Incumbent party vote (share of turnout)			
	(1)	(2)	(3)	(4)
Facebook ads in high saturation × Non-local media outlets	0.001	-0.000	0.001	0.000
	(0.001)	(0.002)	(0.001)	(0.002)
Facebook ads in high saturation × Above-median irregularities × Non-local media outlets	-0.001	0.001		
	(0.002)	(0.003)		
Facebook ads in high saturation × Irregularities Q3 × Non-local media outlets			-0.002	0.001
			(0.003)	(0.002)
Facebook ads in high saturation × Irregularities Q4 × Non-local media outlets			0.005	0.006*
			(0.003)	(0.003)
Facebook ads in low saturation × Non-local media outlets	-0.002	-0.003**	-0.001	-0.002*
	(0.001)	(0.002)	(0.001)	(0.001)
Facebook ads in low saturation × Above-median irregularities × Non-local media outlets	0.002	0.007**		
	(0.002)	(0.003)		
Facebook ads in low saturation × Irregularities Q3 × Non-local media outlets			-0.008***	-0.004
			(0.002)	(0.003)
Facebook ads in low saturation × Irregularities Q4 × Non-local media outlets			0.008**	0.011***
			(0.003)	(0.004)
Spillover in high saturation × Non-local media outlets	0.001	0.000	0.001	0.000
	(0.001)	(0.002)	(0.001)	(0.002)
Spillover in high saturation × Above-median irregularities × Non-local media outlets	-0.002	0.000		
	(0.002)	(0.003)		
Spillover in high saturation × Irregularities Q3 × Non-local media outlets			-0.004	0.000
			(0.003)	(0.002)
Spillover in high saturation × Irregularities Q4 × Non-local media outlets			0.005*	0.006*
			(0.003)	(0.003)
Spillover in low saturation × Non-local media outlets	-0.002	-0.004**	-0.001	-0.003*
	(0.001)	(0.002)	(0.001)	(0.001)
Spillover in low saturation × Above-median irregularities × Non-local media outlets	0.001	0.006*		
	(0.002)	(0.003)		
Spillover in low saturation × Irregularities Q3 × Non-local media outlets			-0.010***	-0.006**
			(0.002)	(0.003)
Spillover in low saturation × Irregularities Q4 × Non-local media outlets			0.008**	0.013***
			(0.003)	(0.004)
Observations	13,234	13,234	13,234	13,234
R^2	0.59	0.71	0.64	0.74
Control outcome mean	0.28	0.28	0.28	0.28
Control outcome std. dev.	0.14	0.14	0.14	0.14
Treatment × covariate interactions		✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. Lower-order interaction terms are included, but omitted to save space. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; (log) amount of FISM funds received per capita; (log) population aged above 18; lagged incumbent party vote share; average years of schooling; share of the population that is illiterate; average number of occupants per room, by household; average number of children per woman; the share of the population with electricity, water, and drainage in their home; the working age share of the population; and the share of households with internet at home. The omitted irregularities category is Q1/2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-sided tests).