Online political information: Facebook ads, electorate saturation, and electoral accountability in Mexico*

José Ramón Enríquez† Horacio Larreguy‡ John Marshall§
Alberto Simpser¶

March 2020

While inexpensive digital technologies like Facebook can spread misinformation, they could also enhance electoral accountability. We study electoral responses to incumbent performance information disseminated through social media and how these responses vary with information campaign *saturation*—the share of an electorate directly targeted with information. We evaluate a non-partisan NGO's campaign that used Facebook ads to inform Mexican citizens about the extent of irregularities in audited government expenditures in their municipality prior to the 2018 general elections. The information campaign was randomized to target 0%, 20%, or 80% of a municipality's electorate. Around 15% of targeted citizens watched at least part of the Facebook video ad. We find that incumbent parties which engaged in negligible irregularities received around 5 percentage points more votes among citizens living in areas directly targeted by the ads. This effect in treated areas was twice as large under 80% as 20% municipal saturation, while the higher saturation municipalities also generated comparable spillovers among non-targeted citizens within the same municipality. Social interactions between citizens, rather than responses by politicians or media outlets, appear to drive both the direct and spillover effects. Information campaign saturation may then help explain the particularly large impacts of information on voting behavior and electoral accountability when it is disseminated by broadcast and digital mass media.

*We thank Daron Acemoglu, David Broockman, Bruce Buena de Mesquita, Chappell Lawson, Craig McIntosh, Elias Papaioannou, Maria Petrova, Vincent Pons, Austin Wright, David Yang, and participants at Chicago's Harris School, the CESS of Oxford in Chile and Oxford, EGAP, Emory, Harvard, ITAM, MIT GOV/LAB, the NYU-CESS Conference, UVA, Wisconsin, and Yale 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 at the AEA RCT Registry and EGAP, and is publicly available at https://www.socialscienceregistry.org/trials/3135 and http://egap.org/registration/4806.

†Department of Government, Harvard University. Email: jrenriquez@g.harvard.edu.
‡Department of Government, Harvard University. Email: hlarreguy@fas.harvard.edu.
§Department of Political Science, Columbia University. Email: jm4401@columbia.edu.
¶Department of Economics and Political Science, ITAM. Email: alberto.simpser@itam.mx.

1 Introduction

Advances in digital communication technologies have created new opportunities for targeting information toward large numbers of citizens at limited cost. These advances have been especially dramatic in the Global South, where the use of internet and social media platforms—which are primarily accessed via cell phones—is rapidly catching up to levels in the Global North and has grown by more than 50% within the last 5 years (Poushter, Bishop and Chwe 2018). The growing availability of these technologies is revolutionizing access to politically-relevant information and democratizing who can provide such information (see Zhuravskaya, Petrova and Enikolopov forthcoming).

While the potential for partisan actors to manipulate or distract citizens with fake news and government propaganda are critical concerns for electoral accountability, 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 are often vulnerable to political capture (Anderson and McLaren 2012; Baron 2006; Besley and Prat 2006), non-partisan actors can enhance voter selection and control of elected representatives (e.g. Barro 1973; Fearon 1999; Ferejohn 1986). This potential for facilitating electoral accountability—the focus of our article—is particularly important in the Global South, where politician malfeasance and low-quality public goods provision remain major challenges (Khemani et al. 2016). However, the evidence that providing voters with information about incumbent performance, via various different technologies, increases electoral accountability is mixed (see Dunning et al. 2019).

A distinctive feature of online communication technologies like Facebook and Twitter, as well as mass broadcast media, is the capacity for information campaigns to reach large numbers of citizens within a given electoral unit. Focusing on the potential benefits of the digital revolution, we argue that information campaign *saturation*—which we define as the share of an electorate with direct access to such a campaign—can induce and amplify the impacts of online information campaigns on electoral accountability by facilitating interactions between citizens.[1] Even among voters with access to information, campaign saturation may: (i) increase the likelihood of belief updating in response to incumbent performance information, by accentuating information diffusion between citizens (e.g. Alatas et al. 2016) or inducing partisan responses and media reporting that in turn increases the probability of voters engaging with the campaign's information; or (ii) coordinate voting on the basis of public signals of incumbent performance in office—rather than competing

[1]Common alternative conceptions of saturation address the number of times that 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 *degree* to which campaign content is absorbed across units within a group (as in Baird et al. 2018).

influences on vote choices, such as clientelism or shared identity—by generating common knowledge (Cornand and Heinemann 2008; Morris and Shin 2002) or facilitating explicit communication or agreements (Chwe 2000; Larson 2017; Little 2016; Shadmehr and Bernhardt 2017) between voters. We thus expect the magnitude of any effect of information campaigns documenting incumbent performance in office on electoral support for the incumbent to increase in the campaign's saturation.

This article leverages a field experiment to estimate the effects of a large-scale online information campaign, and variation in its electorate-level saturation, on electoral accountability at the municipal level during the 2018 Mexican general elections. In particular, we evaluate the impact of a non-partisan campaign by Borde Político—a Mexican NGO that seeks to promote government transparency using digital tools—that used Facebook ads to inform citizens of the amount of federal transfers to municipalities intended for social infrastructure projects benefiting the poor and the share of the municipal incumbent's expenditures under this program that were subject to irregularities. Mexico's independent Federal Auditor's Office (ASF) defines irregularities as funds spent on unauthorized projects or social infrastructure projects that do not benefit the law's intended recipients. Such irregular expenditures often constitute corruption (Chong et al. 2015; Larreguy, Marshall and Snyder 2019). This information was extracted from the ASF's publicly-available audit reports and disseminated via 26-second paid-for video ads 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 left-wing National Regeneration Movement (MORENA) party won by a landslide across federal and local elections.

In collaboration with Borde Político, we randomized whether their Facebook ad campaign targeted 0%, 20%, or 80% of the electorate in the 128 municipalities comprising our sample. Within the treated municipalities, we then randomized the targeting of Facebook ads across multiples of 5 segments (groups of contiguous electoral precincts) in accordance with the 20% and 80% saturation levels. To achieve this, all voting age Facebook users in 1 (4) of every 5 segments within low (high) saturation municipalities were directly targeted with Facebook ads.[2] This randomized saturation design identifies (i) the direct effect of access to the campaign within a given segment, (ii) the indirect—or "spillover"—effect of the campaign in untreated segments of treated municipalities, and (iii) how either segment-level effect varies with municipal saturation (see Baird et al. 2018). 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.[3] Engagement with the campaign was broadly proportionate with the level of access prescribed

[2]Around 70% of Mexicans are Facebook users, so the maximum reach of the ad campaign was closer to 14% and 56% in practice.

[3]Our Facebook analytics data can only distinguish whether users watched the ad at all, for at least 3 seconds,

by the campaign saturation level. A parallel panel survey shows that respondents comprehended and retained the information provided by the video.

Precinct-level electoral returns show that this large-scale digital information campaign significantly affected voting behavior. First, relative to pure control segments, the best-performing incumbent parties—those whose citizens were informed of zero or negligible levels of irregularities—increased their vote share among registered voters in the average segment that was directly targeted by Facebook ads by 4-5 percentage points, or almost half a standard deviation. The vote share of incumbent parties that presided over irregularities in the third quartile of the distribution was unaffected. Incumbent parties in the worst-performing quartile suffered a 1-2 percentage point loss of votes in directly targeted segments, although this was not statistically significant. Sanctioning may have been limited by voters already being informed about poor performance, risk-averse voters becoming less uncertain about the incumbent party's type (see Arias, Larreguy, Marshall and Querubín 2019), or voters already having coordinated on supporting MORENA—the opposition party that eventually won the elections (see Shadmehr and Bernhardt 2017). Since a municipality's level of irregularities is not randomly assigned, we show that our results are robust to adjusting for the interaction between treatment and various covariates as well as different coding of low and high irregularities. While Borde Político's campaign increased turnout by around 1 percentage point across all treated segments, the changes in incumbent party vote share for the best-performing incumbent parties are not driven by aggregate shifts in voter mobilization.

Second, we further demonstrate that these effects in directly targeted segments are largely driven by segments in municipalities that received the higher saturation information campaign. For the least malfeasant incumbent mayors, the incumbent party's vote share increased by 5-6 percentage point in treated segments within the high saturation municipalities, where 80% of the electorate was targeted. In contrast, the 2-3 percentage point increase in the incumbent party's vote share in treated segments within low saturation municipalities, where 20% of the electorate was targeted, was significantly smaller than the effect in treated segments within high saturation municipalities.[4] Since the level of engagement with the Facebook ads was similar across treated segments in the 20% and 80% saturation municipalities, these results imply a strong complementarity between access to Borde Político's Facebook information campaign and a substantial share of other Facebook users within the same municipality also having access to the information provided by the campaign.

We further suggest that the effectiveness of high saturation Facebook ad campaigns is driven by interactions between citizens, rather than substantial persuasion among individuals viewing the ads

for at least 10 seconds, or entirely. Since viewers were informed of the ad's topic area at the outset and the level of irregularities in their municipality was reported just over halfway through the ad, 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.

[4] The average effect across saturation levels is closer to that in high saturation municipalities because there are fewer directly treated segments in low saturation municipalities.

in isolation or politician reactions and/or media reporting induced by Borde Político's ad campaign. Consistent with descriptive data indicating that information diffusion between citizens and explicit and tacit voter coordination are common in our sample, the vote share of the least malfeasant incumbent parties also increased by around 5 percentage points in untreated segments within high saturation municipalities, whereas we fail to detect such spillover effects in untreated segments within low saturation municipalities registering similarly negligible levels of malfeasance. This suggests that social interactions induced by the campaign account for most of the effect in directly-treated segments.[5] In contrast, we find no evidence to suggest that other potential amplification mechanisms—specifically, online political responses or media reporting—could account for the results. Our design does not allow us to distinguish whether changes in voting behavior reflect individual belief updating induced by information sharing between citizens or voter coordination (or both). Nevertheless, our findings indicate that subsequent social interactions are key mechanisms in enabling high saturation campaigns to amplify the effects of mass online information campaigns.

This study makes several main contributions. First, we highlight the substantial potential for electoral impact of information disseminated via social media platforms during election campaigns. Our evaluation of Borde Político's Facebook ad campaign suggests that social media may play a greater role in persuading citizens to change who to vote for than in convincing citizens to turn out. Whereas we observe small increases in turnout comparable to Facebook's own "get out the vote" campaigns (Bond et al. 2012; Jones et al. 2017), the provision of non-partisan information pertaining to incumbent performance shows that mass campaigns can far more substantially influence vote choice. These large effects align with recent evidence that Facebook polarizes political attitudes in the U.S. (Allcott et al. 2020; Levy 2019) and that micro-targeted ads significantly increased self-reported support for Donald Trump in the 2016 U.S. presidential election (Liberini et al. 2018), although Rink (2019) detects only small effects of partisan campaign ads disseminated via Facebook and Google in Germany. Our results also align with a non-experimental literature examining the effects of the internet more generally, which finds growing political impacts of the internet over time (Campante, Durante and Sobbrio 2017) and that access to the internet helps citizens to hold their governments to account (Miner 2015). However, by experimentally manipulating ad targeting across Facebook users, our design overcomes the difficulty of distinguishing the effects of social media content from social media penetration that Zhuravskaya, Petrova and Enikolopov (forthcoming) highlight with regard to such observational studies.

Second, our findings counterbalance concerns that fake news disseminated via social media may have shaped vote choices in the U.S. (e.g. Allcott and Gentzkow 2017) as well as Brazil, India, and Nigeria.[6] In contrast, we provide the first evidence that factual and non-partisan information

[5] We find no indication that the substantial indirect effects reflect inaccuracy in Facebook's spatial targeting of ads.

[6] India's 2019 election was even described as its "WhatsApp election" by the *Financial Times* ("India: the WhatsApp election," May 5, 2019).

disseminated via Facebook's low-cost ads can promote electoral accountability. Like radio-based anti-vote buying campaigns that reduced support for India's more clientelistic parties (Vasudevan 2019) and election-related information on the radio that increased electoral competition in the U.S. (Panagopoulos and Green 2008), the increased support that we observe for incumbent parties that are ostensibly less corrupt will most likely increase citizen welfare in Mexico. In light of growing efforts to regulate social media during elections in response to widely-circulating fake news (e.g. in India and Turkey), our findings imply that democratizing the control of content provision can also support electoral accountability.

Third, we demonstrate that the impact of incumbent performance information disseminated via social media is, to a substantial degree, causally driven by the saturation of the information campaign at the electorate level. While Adida et al. (2019) also experimentally varied the saturation of an accountability campaign at the electorate level in Benin, their greatest level of saturated targeting (15%) is substantially lower than in Borde Político's campaign.[7] At this lower level of saturation, they find that saturation principally amplified the effect of civics training, rather than the effect of the incumbent performance information provided alongside such training. The stark differences that we find across the 20% and 80% information campaign saturation levels more broadly suggest that saturation could account for the notable heterogeneity in the treatment effects of disseminating incumbent performance information. Indeed, low saturation information campaigns that reached few other voters have had little impact on treated individuals' vote choices (Adida et al. 2019; Boas, Hidalgo and Melo 2019; Dunning et al. 2019; Humphreys and Weinstein 2012; Lierl and Holmlund forthcoming). Conversely, information that became accessible to large shares of the electorate, via the media (e.g. Banerjee et al. 2011; Ferraz and Finan 2008; Larreguy, Marshall and Snyder 2019; Marshall 2019*b*) or concentrated leafleting (Arias, Larreguy, Marshall and Querubín 2019), has generated greater electoral rewards and punishment on the basis of incumbent performance in office.[8] Our results thus provide causal evidence that saturation—a defining characteristic of broadcast, print, and now digital media—may be a key driver of the larger impacts of information delivered by mass media.

Fourth, in providing evidence suggesting that interactions between citizens drive saturation's effects, we further highlight how social interactions can amplify information's effects on electoral accountability. This finding complements recent evidence suggesting that information campaigns have diffused within communities or households to influence vote choices (Bhandari, Larreguy and

[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 we focus on and—perhaps unsurprisingly—fail to detect differential saturation effects.

[8] Appendix section A.1 describes how prior studies vary in terms of information campaign saturation. However, 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 (2020), where the information diffusion that resulted from a very low scale campaign was substantial in rural Senegal.

Marshall 2020) and turnout (Fafchamps, Vaz and Vicente Forthcoming; Nickerson 2008), coordinated voters around better candidates (Arias, Balán, Larreguy, Marshall and Querubín 2019), and produced larger effects when widespread campaigns were common knowledge (George, Gupta and Neggers 2019). Beyond elections, communication between citizens also appears to have stimulated protest and collective action (e.g. Acemoglu, Hassan and Tahoun 2017; Enikolopov, Petrova and Makarin 2019; Fergusson and Molina 2019; García-Jimeno, Iglesias and Yildirim 2018; Manacorda and Tesei forthcoming; Pierskalla and Hollenbach 2013; Steinert-Threlkeld 2017). In contrast with these studies, our randomized saturation design enables us to identify spillover and saturation effects to show that social effects are particularly prominent in high saturation campaigns. Our findings thus align with recent evidence of complementarities in protest participation between students in Hong Kong (Bursztyn et al. 2019).

Finally, our finding that a non-partisan NGO campaign can influence vote choices relates to a broader partisan persuasion literature. Across various contexts, partisan campaign ads have proven effective at winning votes (Da Silveira and De Mello 2011; Gerber et al. 2011; Larreguy, Marshall and Snyder 2018; Rink 2019; Spenkuch and Toniatti 2018) and debates between candidates have increased support for the best-performing candidates (Bidwell, Casey and Glennerster forthcoming; Bowles and Larreguy 2019; Platas Izama and Raffler 2019). Similarly, news content has persuaded voters to switch parties (Adena et al. 2015; DellaVigna and Kaplan 2007; DellaVigna et al. 2014; Enikolopov, Petrova and Zhuravskaya 2011; Gentzkow, Shapiro and Sinkinson 2011; Martin and Yurukoglu 2017). We observe larger reduced form effects of non-partisan independent audit information than most of these studies do for partisan mass media, suggesting that maintaining a role for transparency-oriented NGOs may be critical in improving electoral accountability in the Global South.

This article is structured as follows. Section 2 discusses potential mechanisms for a saturation effect. Sections 3 and 4 describe the context and experimental evaluation of Borde Político's Facebook ad campaign. Section 5 then describes the campaign's reach, before we report our main results in section 6 and explore mechanisms in section 7. Section 8 concludes.

2 How information campaign saturation could affect electoral accountability

Theories of electoral accountability posit that information about an incumbent's performance in office can help citizens to select and control elected representatives. First, incumbent performance information can mitigate adverse selection problems by helping voters to identify politicians that are likely to perform competently or pursue policies aligned with their interests in the future (Fearon

1999; Rogoff 1990). Second, incumbent performance information can help voters to replace shirking or corrupt politicians, which may in turn reduce incentives for future incumbents to engage in rent-seeking behavior (Barro 1973; Ferejohn 1986). These theories predict that providing relevant incumbent performance information will induce voters to sanction poorly-performing incumbents and reward highly-performing politicians at the ballot box, especially where performance indicators deviate most from voters' prior expectations (see Arias, Larreguy, Marshall and Querubín 2019; Kendall, Nannicini and Trebbi 2015).

In light of the mixed evidence that information campaigns help voters to reward (sanction) better (worse) performing incumbents,[9] and the growing degree to which political information disseminated through social media is reaching substantial shares of the electorate, we further consider the degree to which the *saturation* of an information campaign influences electoral accountability. We argue that the proportion of eligible voters within a given electoral unit that can access incumbent performance information could induce or amplify the effects of such information on electoral accountability among voters targeted by an information campaign through two primary mechanisms: information diffusion and voter coordination. While the former mechanism may generate a complementarity between information provision and a campaign's saturation by increasing the likelihood that citizens update their beliefs in response to the information campaign, the latter might also do so by increasing the likelihood of coordinated voting behavior for any given posterior belief about the incumbent's type or effort.

Most straightforwardly, saturation could amplify an information campaign's effect among those with access to the campaign by increasing the probability that information ultimately reaches—and is internalized by—its targets. While voters often ignore or only cursorily view pamphlets, text messages, broadcast media programming, or online content providing unsolicited politically-relevant information (e.g. Dunning et al. 2019), saturation could increase engagement through at least two channels. First, where citizens discuss political issues with each other, campaign saturation is likely to increase exposure to information about incumbent performance. This could result from citizens directly sharing their information with others within their social network (e.g. Alatas et al. 2016; Alt et al. 2020; Bhandari, Larreguy and Marshall 2020; Buntaine et al. 2018; García-Jimeno, Iglesias and Yildirim 2018) or encouraging others to seek out political information (Marshall 2019*a*). Second, high saturation information campaigns could increase the likelihood that parties and media outlets become aware of the information or learn that it is of interest to their audience. Parties or media outlets may then retransmit the information to large audiences, and thereby facilitate greater belief updating. Consequently, to the extent to which an information campaign disseminates relevant content capable of influencing a citizen's capacity to hold politicians accountable, greater saturation is expected to amplify the campaign's effects on both voters that

[9] See Appendix section A.1 for a detailed review of extant studies.

were directly targeted by the campaign and voters connected to those targeted by the campaign by increasing the probability of engaging with the information.

High saturation information provision could also induce or amplify voter responses by coordinating voting behavior through at least two channels. The first *tacit* coordination channel relies on common knowledge, such that—even without explicit communication between citizens—a high saturation information campaign may lead voters to believe that many other voters also received the same information. Where individual actions are complements, the global games framework of Morris and Shin (2002) predicts that information campaigns that represent public signals can coordinate individuals around the action corresponding to the realization of the public signal. Vote choices may be strategic complements in our electoral setting if bloc voting more effectively signals approval or disapproval of incumbent performance or if voters obtain expressive utility from aligning their votes with their peers. Following Morris and Shin (2002), such motivations would lead voters to downweight their prior beliefs in favor of tacitly coordinating their votes on incumbent parties publicly revealed to have performed well in their municipality. Consistent with such a role for common knowledge, George, Gupta and Neggers (2019) find that Indian voters informed via SMS that a candidate has been accused of serious crimes are more likely to vote against the candidate when they are also informed that many other citizens were sent the same SMS. Cornand and Heinemann (2008) further prove that responsiveness to the performance revealed by a public signal among the citizens that receive the signal increases with the share of citizens with access to the public signal. The common knowledge rationale therefore also implies that information campaign saturation should amplify any positive (negative) effect on incumbent party vote share of a public signal indicating that incumbent performance exceeded (fell below) prior expectations.

The second, and related, channel involves higher levels of saturation increasing the probability of *explicit* coordination. Explicit coordination involves direct communication between citizens that facilitates implicit understandings or explicit agreements to synchronize their vote choice. This could emerge in response to updating based on the information's content or could be sparked simply because providing information increases the salience of malfeasance in voters' decision calculus. Shadmehr and Bernhardt (2017) consider a model of the former type, where communication between citizens both diffuses information about the payoff of collective action and prevents coordination failure. Applied to our setting, such communication will increase support for the incumbent party about which citizens update, relative to a setting where citizens receive equivalent signals in isolation, when the challenger party is relatively appealing *ex ante*. This is because, by pooling signals, communication conveys the more accurate information required to overcome the initial preference for the challenger, as in the 2018 Mexican context where MORENA won a landslide victory despite holding very few municipal offices. Other models similarly suggest that communication within networks can induce coordinated citizen behaviors, such as prioritizing political accountability over clientelistic equilibria, that citizens already regarded as socially optimal (e.g. Chwe 2000; Larson 2017; Little 2016; Morris and Shin 1998). In line with these predictions, Arias, Balán, Larreguy, Marshall and Querubín (2019) find that an information campaign in Mexico induced explicit and tacit coordination that influenced vote choices. Beyond voting behavior, Acemoglu, Hassan and Tahoun (2017) and Lynch (2011) similarly argue that social media facilitated protests during the Arab Spring, while Enikolopov, Petrova and Makarin (2019) and Fergusson and Molina (2019) respectively provide more concrete evidence of this in Russia and across the globe. Manacorda and Tesei (forthcoming) find that cell phones can play a similar role in Africa.

Taken together, the preceding information diffusion and voter coordination mechanisms generally predict that any effect of information dissemination on electoral accountability is likely to be induced or amplified by greater campaign saturation. We test this hypothesis in the context of a large-scale information campaign undertaken before Mexico's 2018 elections.

3 Mayoral malfeasance and accountability in Mexico

Mexico's c. 2,500 municipal governments are led by mayors typically elected to three-year terms, which became renewable for the first time 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 accountability remains limited, and corruption is still common.

3.1 Independent audits of municipal spending

A key 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 are mandated exclusively for 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.[10] 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.

The use of FISM transfers is audited in around 200 municipalities each year by Mexico's independent Federal Auditor's Office (ASF). ASF audits are announced after spending has occurred, and address the spending, accounting, and management of FISM funds from the previous fiscal year. Municipalities are selected by the ASF on the basis of the importance of FISM transfers to

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

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 (including concurrent federal audits of other programs) (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 the expenditure of FISM resources. Irregularities typically entail funds that were spent on projects not benefiting the poor (based on the distribution criteria above) or spent on unauthorized projects that did not constitute social infrastructure projects (e.g. personal expenses and election campaigns). The audit reports indicate that such irregularities typically arise from failing to demonstrate that the project benefited its intended recipients, the transfer of funds to non-FISM bank accounts or contractors, or failures to produce documentation proving that expenses related to claimed projects. These actions often reflect corruption in the form of kickbacks, preferential contracting, and embezzlement. Between 2009 and 2018, the ASF determined that 17% of funds spent were subject to irregularities.[11]

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 where there is a large audience for such information (Larreguy, Marshall and Snyder 2019). However, because media coverage is not widespread and voter engagement with news programming varies, further dissemination of such information can significantly alter voters' beliefs and voting behavior. Indeed, Arias, Larreguy, Marshall and Querubín (2019) find that distributing the results of ASF reports via nonpartisan 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 can breed voter disengagement, with a particularly detrimental effect on support for challenger parties. This article complements these prior studies by investigating how the provision of information via social media and the extent of its saturation facilitate electoral accountability.

[11] Given that other programs and non-federal transfers are not subject to such audits, mayoral malfeasance could be greater on other dimensions. Nevertheless, we expect malfeasance across areas to be positively correlated, and thus that information about irregularities in FISM expenditures—which represent a substantial share of a municipality's budget—will be indicative of an administration's broader malfeasance.

3.2 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 populist PRI competed against either the relatively urban right-wing National Action Party (PAN) or the PRI's more rural 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 and often displaced the PRD. Although MORENA's local presence was initially limited, it swept the 2018 elections as López Obrador's message of change won him the presidency by a landslide. MORENA's national success carried over to local elections as well, with MORENA claiming multiple governorships and hundreds of mayoral offices across the country. The 2018 election was thus unusual in the extent of emphasis on reducing corruption and the overwhelming success of a party that had previously held few legislative or executives offices.

Municipal election campaigns in Mexico are generally oriented around political parties, rather than specific candidates, for several reasons. First, given that consecutive re-election for mayors was only permitted for the first time in 2018, citizens are generally much better informed about parties than individual politicians (e.g. Arias, Larreguy, Marshall and Querubín 2019; Chong et al. 2015; Larreguy, Marshall and Snyder 2019). Second, voters may recognize that Mexico's main parties use distinct candidate selection mechanisms that select candidates with similar characteristics over time (Langston 2003). Consequently, voters have held parties responsible for the actions of individual politicians (e.g. Chong et al. 2015; De La O 2013; Larreguy, Marshall and Snyder 2019; Marshall 2019*b*). Despite the fact that only 22% of mayors sought re-election in 2018, there are thus good reasons to believe that voters will make inferences about the party of the mayor whose audited expenditures are publicized and vote accordingly.

3.3 Political information and 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, 72% of adults own a smartphone—the primary means through which adults access the internet in Mexico—and 64% of adults used social media in 2018; 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 ranks 5th in the world in terms of active Facebook users. WhatsApp has become the messaging service of choice, and is the most used cell phone app.

Growing access to digital information has emerged alongside substantial amounts of credible and fake political information disseminated through social and traditional media. Fake news was a particular concern during the 2018 election campaign, where political parties were accused of disseminating fake news aided by bots to influence voter behavior.[12] The attacks were largely directed against the eventual winner López Obrador.[13] However, many other candidates across all races were also affected by similar types of attacks.[14] Due to their popularity among Mexican citizens, Facebook and WhatsApp were the prime channels for spreading real and fake news in the form of videos, images, and memes. Several Facebook pages that were identified as the most prolific fake news distributors had between one and two million followers around the election.

4 Research Design

Our study estimates whether non-partisan incumbent performance information disseminated via Facebook ads affects electoral accountability, and the extent to which the effect among voters targeted by the ads is moderated by an information campaign's saturation. We partnered with Borde Político—an NGO primarily based in Mexico City, which uses digital technologies to promote government transparency across the country—to evaluate the impact of their online accountability campaign ahead of the July 1, 2018 elections. Borde Político's Facebook ad campaign, which focused on the municipal elections, provided voting age adults with information about the FISM program and the share of audited resources that the ASF found to be subject to irregularities. Widespread access to social media enabled the study to randomly vary the share of the municipal population that was targeted with this information via Facebook. This section describes the treatment conditions, sample, experimental design, measurement of outcomes, and estimation, and concludes by discussing ethical considerations.

[12]For example, due to the way that Facebook's algorithm works, "likes" of a Facebook page or ad increase their visibility. Facebook pages criticizing López Obrador featured posts with thousands of "likes," but no other reactions or comments, suggesting the work of bots. See here for more details.

[13]For example, fake news articles that claimed that López Obrador's wife posted on Twitter that she was disgusted by indigenous people—in a country where official figures indicate that 21.5% of the population is indigenous—were widely shared on Facebook (see here for more details). A fake poll in a major national newspaper suggesting that the PAN candidate was within 5 percentage points of López Obrador's—the latter eventually beat the former by more than 30 percentage points—was also widely circulated by PAN candidates over social media (see here for more details). Fake pictures of rallies with very few attendees were circulated to claim that López Obrador's support was deflating (see here for more details).

[14]See, for example, here, here, and here.

4.1 Treatment conditions

Like earlier studies that provided citizens with incumbent performance information, Borde Político's information campaign reported the results of the ASF's audit in a given municipality.[15] Citizens that received the information were first informed that the FISM program transfers federal funds to municipalities for social infrastructure projects benefiting the poor. They were then informed of how much money their municipal government received, and the percentage of the audited funds that were subject to irregularities in terms of violating FISM spending regulations.[16]

As part of Borde Político's broader transparency campaigns, this information was disseminated to Facebook users via a Facebook video ad. Figure 1 shows the slides that make up the 26-second video. The first slide was designed—based on initial pilots—to attract viewers, while the share of a municipality's FISM expenditures that were subject to irregularities was reported in the 17th second. To bolster credibility, 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 also click to access the municipality-specific Facebook page that promoted the ad.[17] These pages included a cover photo highlighting the FISM funds received 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). Each municipality ad campaign ran for a week, concluding on June 27, 2018—the last day of official campaigning. Incumbents thus had no time to meaningfully alter their performance in office before the election in response to the ads, and parties had little time to respond during the campaign.

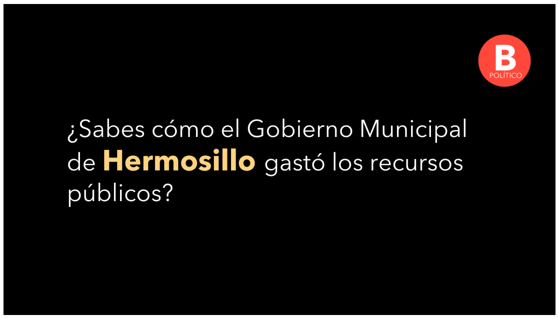
We randomly varied the saturation of Borde Político's Facebook ad campaign across municipalities. Accordingly, Facebook ads were geographically targeted with the capacity to reach 20% of Facebook users of voting age (18+) in low saturation municipalities, while Facebook ads sought to reach 80% of Facebook users of voting age in high saturation municipalities.[18] The 20% and 80% saturation levels were chosen to capture a meaningful difference in saturation that could plausibly alter levels of information diffusion or coordination, while also maximizing our statistical power

[15] While the information content provided is similar to prior interventions in Mexico (e.g. Arias, Larreguy, Marshall and Querubín 2019; Chong et al. 2015), this study differs by focusing on the impact of digital dissemination and municipal campaign saturation. By leveraging similar information content, prior studies help benchmark these effects.

[16] Information was not cross-sectionally or temporally benchmarked because prior studies detect no effect of additionally providing information from comparable Mexican municipalities (Arias et al. 2018) and because some municipalities had not previously experienced an audit.

[17] A separate page was created for each basic and common knowledge version of the video in every municipality (see below).

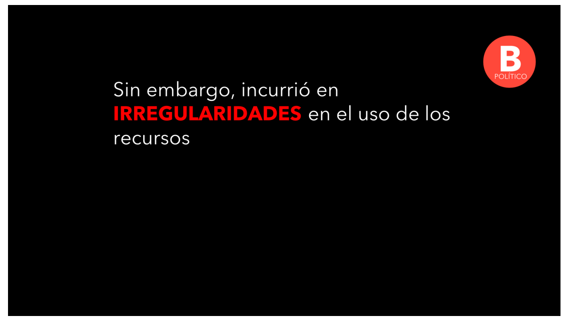
[18] In all treated (and some control) municipalities, individual WhatsApp messages were sent to a mean of 50 surveyed registered voters as part of a concurrent panel survey designed to understand the mechanisms underlying the Facebook campaign. Since this number represents a negligible fraction of the municipal population, we disregard them when defining municipal treatments. This approach is supported by the lack of a significant difference in electoral outcomes across control municipalities that did and did not contain respondents that received WhatsApp messages.



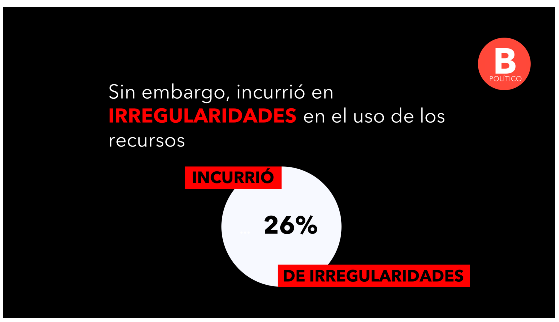
(a) Slide 1 (4 seconds)

(b) Slide 2 (4 seconds)

(c) Slide 3 (5 seconds)

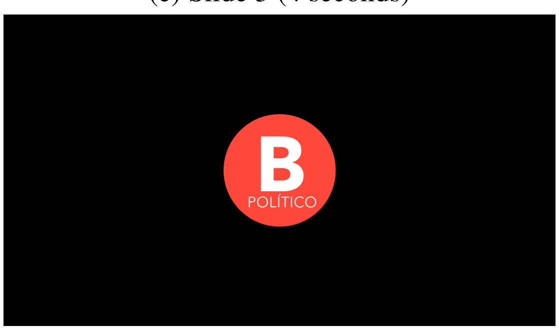


(d) Slide 4 (4 seconds)



(e) Slide 5 (4 seconds)

(f) Slide 6 (4 seconds)



(g) Slide 7 (1 second)

Figure 1: Example of the slides included in the ad video (from Hermosillo, Sonora)

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 "Received $65 million for infrastructure;" slide 4 says "However, it incurred in irregularities in the spending of the funds;" slide 5 adds "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."

to estimate direct, indirect, and differential saturation effects (see below).[19] The average municipal ad campaign cost around US$200, representing a small fraction of a typical municipal election campaign's budget.

Borde Político's Facebook ad campaigns could not be designed to ensure that ads would reach all voting age adult Facebook users within targeted locations a certain number of times. Rather, for a given investment, Facebook allows the purchaser of ads to specify the maximum possible reach of a campaign (in terms of the number of users) within a geographic area for a particular demographic. Consequently, Borde Político's ad campaigns were funded to be able to reach the designated 20% or 80% of voting age individuals in low and high saturation municipalities as many times as possible and at equal rates across municipalities. While Facebook does not publicly disclose its constantly-evolving technology used to identify user locations, our conversations with Facebook staff indicate that whether a given Facebook user is targeted by a geographically-constrained ad depends primarily on the location that Facebook identifies a user as spending most time in, based on user-specific GPS data. For most users, this is their home.[20] When a user's GPS data is unavailable, targeting is based on data including the user's IP address, search traffic, and the locations of a user's friends. Since 88% of users accessed the ads via a mobile device, ads are generally likely to be targeted with a high degree of accuracy. We provide evidence consistent with this in our discussion of mechanisms below.

The ad's content was also subtly randomized to explicitly vary common knowledge about the ad campaign's reach. As in George, Gupta and Neggers (2019), Facebook users in some locations were informed that the ad campaign could reach 20% or 80% of citizens in their municipality. This entailed adding the slide shown in Appendix Figure A5 to the end of the video.[21] We ultimately observe no discernible differences in viewership of—or reactions to—the ads with and without explicit common knowledge communication (see Appendix Table A10), while this variant of the treatment also did not differentially affect voting behavior (see Appendix Table A11). This is likely because few viewers reached the end of the ad (see Table 1). For these reasons, we henceforth pool the Facebook ads with and without common knowledge in all analyses.

[19] Following Baird et al. (2018), we minimized the equally-weighted sum of the standard errors for treatment and spillover effects, where municipalities were equally split between the control, low saturation, and high saturation conditions.

[20] The geographic areas covered by Borde Político's Facebook ads (defined below) are large enough that they will often encompass both the home and workplace of Facebook users.

[21] To avoid deception, this information always reflected the true share of Facebook users in the municipality that were targeted by the campaign.

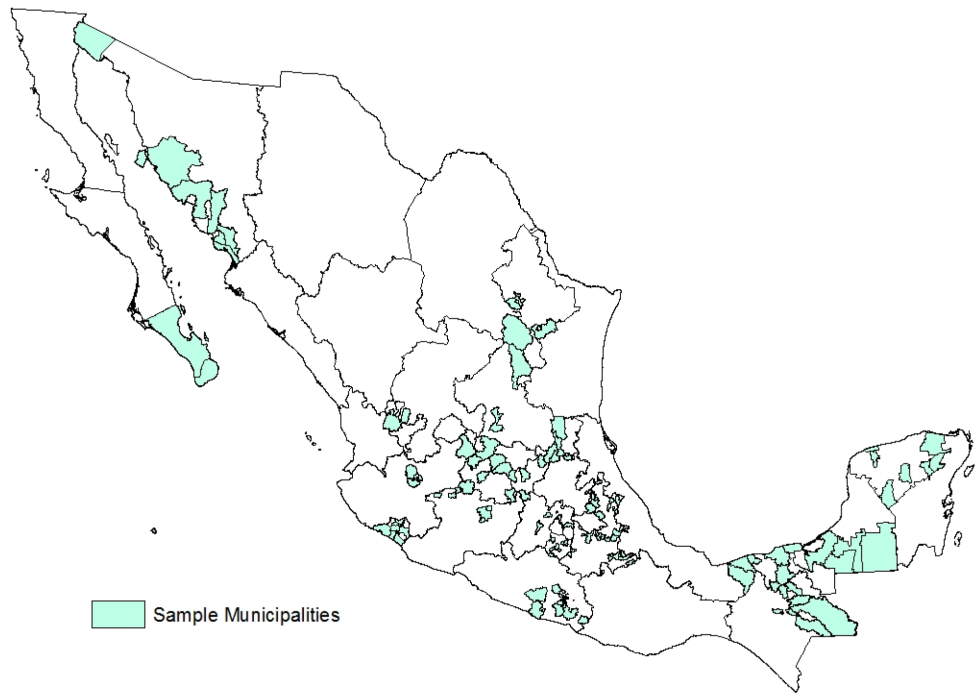


Figure 2: The 128 municipalities included in our sample

4.2 Sample of municipalities

Across 2017 and 2018, the ASF released audit reports pertaining to FISM expenditures in 561 municipalities.[22] Of these, 128 municipalities satisfied our conditions 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.[23] The resulting sample of eligible municipalities is shown in Figure 2. The sample collectively contains around 30 million people, roughly a quarter of Mexico's population, and is broadly nationally representative.

Figure 3 shows that the majority of ASF audits in these municipalities reported irregularities between 0% and 10%. Exactly zero irregular spending was found in 61 of our 128 municipalities. The mean share of irregular spending across municipalities was 9.2%, with a positive skew driven

[22]The two delegaciones in Mexico City were excluded because such delegaciones operated differently from municipalities during the relevant time period.

[23]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. The other 7 states did not hold municipal elections.

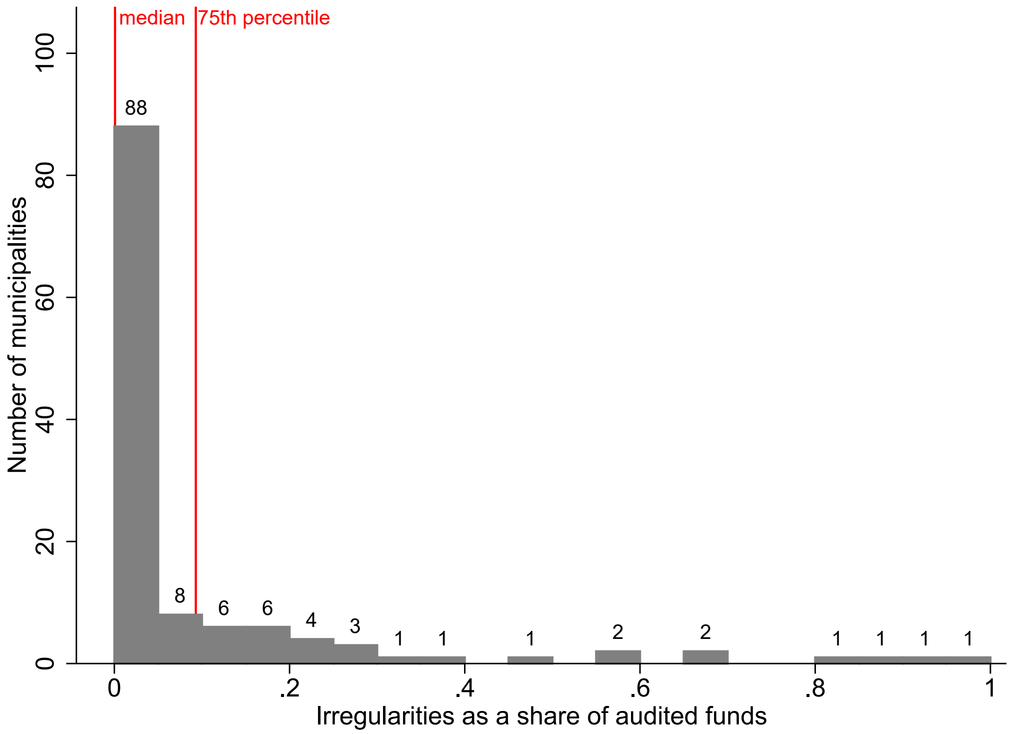


Figure 3: Distribution of irregularities across municipalities in our sample

Notes: Each band has a width corresponding to a 0.05 share. The 25th, 50th, and 75th percentiles of the distribution of irregularities are, respectively, 0%, 0.07%, and 9.3%. The mean is 9.4%, and 61 municipalities registered exactly 0% irregularities.

by several egregious cases. Given the low expectations of politicians recently documented in Mexico (Arias, Larreguy, Marshall and Querubín 2019), the ads likely reported better performance than most citizens expected.

4.3 Experimental design

4.3.1 Aggregate-level treatment assignment

To estimate the electoral effects of Borde Político's Facebook ad campaign, we designed a two-level randomization strategy. This first assigned campaign saturation at the municipality level and then assigned Facebook ads to segments (defined below) within municipalities selected to receive a non-zero campaign saturation. This design enables us to identify the effect of information provision on voting behavior in treated segments, the extent to which information provision affected non-treated segments within partially treated municipalities, and whether these effects vary by municipal saturation level.

The saturation of Borde Político's Facebook ad campaign was randomized at the municipal level as follows. 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, with the exception of 2 rump municipalities that formed an additional block governed by different parties. For simplicity, we henceforth exclude the rump block from our analysis.[24] 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 targeted to be able to reach 20% of adults within the municipality; and
3. *High saturation information campaign*: Facebook ads were targeted to be able to reach 80% of adults within a 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 A3 shows that campaign saturation is well balanced across predetermined municipal-level covariates.

We then randomized the targeting of Facebook ads to geographic areas within treated municipalities. To be able to reach up to 20% and 80% of adults in low and high saturation cases, we divided each municipality into (multiples of) 5 equally-populated "segments." In small 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 defined by contiguous electoral precincts—Mexico's smallest geographical electoral unit—that form compact polygons with similar populations of individuals aged 18 or above (according to the 2010 Census).[25] Complete randomization was used to assign one in five segments within low saturation municipalities, and our in five segments within high saturation municipalities, to be targeted by Borde Político's Facebook ads.[26] As Appendix Table A4 shows, the segment-level treatment conditions are also well balanced across predetermined covariates.

[24]This deviation from our pre-analysis plan was deployed primarily to simplify estimation by maintaining a constant probability of treatment assignment across municipalities (see Appendix section A.6). Appendix Table A2 reports similar results when the two small municipalities in this residual block are included and differential probabilities of treatment assignment are accounted for.

[25]These segments were generated by the freely-downloadable redistricting program Auto-Redistrict (autoredistrict.org). This software allows users to redistrict blocks of precincts into "districts" to maximize the contiguity, compactness, and equal population of districts. Precinct allocation was then manually adjusted at the margins to smooth edges in order to facilitate ease of targeting with Facebook ads (given the targeting constraint of needing to pick points with 1km radii). The latter adjustment effectively slightly relaxed the population equality constraint. The total number of segments is not a multiple of 5 because one small municipality contains only 3 electoral precincts.

[26]Whether a treated segment would additionally receive common knowledge ads informing targeted citizens of how many other citizens within their municipality also had access to the ad was also randomized. Within the large majority of low-saturation municipalities with only five segments, only one segment was treated; complete randomization deter-

4.3.2 Individual-level treatment delivered via WhatsApp

A parallel panel survey was conducted online and by telephone to help illuminate any effects of the ad campaign on voting behavior. This survey yielded complete baseline and endline responses from around 2,000 registered voters that use WhatsApp within the 128 municipalities.[27] The 20-minute baseline survey was conducted over 3 weeks in early June 2018, while the 20-minute endline survey was conducted over the month after the election. Since the sample that ultimately completed the endline survey is unlikely to be representative of the broader electorate, we extrapolate from the panel survey to our sample of municipalities with caution.

As an additional randomized treatment, the Facebook ad video and the information accompanying the ad on Facebook, were sent via a WhatsApp message in the week preceding the election to 80% of baseline respondents within both treated municipalities and 23 of the pure control municipalities (see Appendix section A.4). An example of the WhatsApp message is shown in Appendix Figure A2. Treated survey respondents also received a similar followup message that included the infographic shown in Appendix Figure A1.

Unfortunately, we encountered significant differences in the endline response rates of different types of baseline survey respondents across municipalities assigned to different Facebook ad saturation levels (see Appendix Table A5). We are thus unable to leverage variation in saturation to study the effect of municipal Facebook ad saturation on survey outcomes. However, this problem does not arise for the individual-level WhatsApp treatment among the c.1,500 treated and control individuals in the municipalities where Facebook ads were delivered (see Appendix Table A6). We therefore restrict our analyses of survey-level outcomes to examining within-municipality random variation in being sent Borde Político's Facebook ad video via an individual WhatsApp message.

4.4 Measurement of outcomes

Our main outcomes are taken from the precinct-level electoral returns collated by Mexico's state electoral institutes. We focus primarily on the municipal incumbent party's vote share, using the

mined whether that segment received the common knowledge treatment. In the few low saturation municipalities with a multiple of five segments, receiving an equal number of non-common knowledge and common knowledge treated segments was prioritized. Within high saturation municipalities, half of the treated segments (i.e. 40% of the municipality's segments) received the common knowledge information and the other half did not. As noted above, we focus on the results from analyses that pool ads with and without common knowledge.

[27]The survey was conducted by GeoPoll. They generated a sample based on calling and messaging randomly-generated cell numbers (based on areas codes local to our municipalities), Telmex landline numbers, and the completion of an online Qualtrics survey recruited via a separate Facebook ad campaign solely seeking to recruit survey respondents (which made no reference to Borde Político's Facebook ad campaign). We aimed to recruit 31 respondents per municipality for a baseline survey, and 20 of these for endline survey. Recruitment rates differed across municipalities, but yielded around 14 endline respondents in the average municipality. To incentivize continued participation, respondents that completed the baseline survey were entered into a lottery to win one of 10 prizes with a value equivalent to a new smartphone; an additional independent lottery was used for the endline survey as well.

predetermined number of registered voters as the denominator. Since we find a limited impact on turnout, we obtain similar results when the vote share denominator is the total number of votes cast (see Appendix Tables A13 and A14). Our focus on municipal incumbent parties as the unit of analysis accords with prior studies in Mexico's party-centric electoral context (e.g. Arias, Larreguy, Marshall and Querubín 2019; Chong et al. 2015; Larreguy, Marshall and Snyder 2019). Electoral returns were available for 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. Combined with our exclusion of the block with two rump municipalities, this yielded a final sample of 124 municipalities.

To measure engagement with the treatments and possible intermediary mechanisms, we augment our administrative electoral data with two additional data sources. First, Facebook analytics data associated with each municipality's ad campaigns allows us to measure, at the municipal level, how many and what types of people saw the ads (and for how long) as well as gauge the extent of user interactions. Second, our endline survey also elicited engagement with the WhatsApp messages, respondents' perceptions of the source and credibility of the messages, and various types of beliefs.

4.5 Estimation

Our pre-registered specifications leverage the multiple layers of randomization to identify the effects of Borde Político's Facebook ad campaign on precinct-level electoral results.[28] First, we follow Baird et al. (2018) in leveraging the segment-level ad targeting to estimate direct and indirect average treatment effects on precinct-level outcomes using the following specification:

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

where Y_{psm} is an outcome in precinct p within segment s of municipality m, Y_{psm}^{lag} is a lag of the outcome, $Facebook\ ads_{sm}$ is an indicator for a treated segment s receiving Facebook ads, $Spillover_{sm}$ is an indicator for an untreated segment s located within a treated municipality, and μ_b are fixed effects for the blocks of three similar municipalities within which treatment was assigned. The reference category is the set of precincts from control municipalities assigned to receive no Facebook ads.[29] Observations are weighted by the design's inverse probabilities of treatment assignment and, to weight segments equally, each precinct's share of the segment's 2010 adult population aged 18 or above. Standard errors are clustered by municipality for electoral outcomes throughout because

[28]Minor deviations from the pre-analysis plan are explained in Appendix section A.6. Additional pre-registered specifications that are not reported in the main article due to space constraints are reported in Appendix section A.6.

[29]This includes zero-saturation municipalities where only 0.02% of registered voters received WhatsApp treatments on average.

saturation was randomized at the municipal level.

The coefficients β and γ in equation (1) respectively capture the direct intent to treat effect of being targeted by the Facebook ad campaign and the indirect spillover effect of being located in a non-targeted segment within a treated municipality, both relative to the pure control condition. Since our theoretical expectations are conditional on the content of the information provided by the Facebook ads, we did not pre-register a directional hypothesis for the *average* effects within the entire sample.

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

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

where the differences $\beta_2 - \beta_1$ and $\gamma_2 - \gamma_1$ respectively capture the differential effect of the direct and indirect Facebook ad treatments attributable to the high rather than the low saturation campaign. As argued above, we expect $\beta_2 - \beta_1 > (<) 0$ and $\gamma_2 - \gamma_1 > (<) 0$ when the average treatment effect is positive (negative). We pre-registered one-sided tests of these expectations that campaign saturation induces or amplifies the effects of information provision.

As with all informational interventions, it was not obvious *a priori* how information content would affect support for incumbent parties. Increases or decreases in support are likely to depend on whether reported irregularities exceed voters' prior expectations and whether/how different reported irregularities coordinate voter behavior. Since we could not systematically measure prior beliefs across all segments, we follow Ferraz and Finan (2008) and others in generating bins where reported irregularities are likely to exceed and fall below expectations. We divide the distribution of reported irregularities into quartiles (see also Larreguy, Marshall and Snyder 2019). The bottom two quartiles are pooled because more than 25% of municipalities registered zero irregularities. Irregularities in the third quartile range from 0.08% to 7.4% of audited FISM funds, with a mean of 2.2%. Irregularities in the top quartile range from 7.4% to 100%, with a mean of 31.7%. The interactive specifications then extend equations (1) and (2) by including interactions between treatment conditions and a municipality's irregularities quartile. For example, the interactive version of equation (1) entails estimating:

$$\begin{aligned} Y_{psm} \quad = \quad & \alpha Y_{psm}^{lag} + \beta_1 Facebook\ ads_{sm} + \beta_2 \left(Facebook\ ads_{sm} \times Q3\right) \\ & + \beta_3 \left(Facebook\ ads_{sm} \times Q4\right) + \gamma_1 Spillover_{sm} + \gamma_2 \left(Spillover_{sm} \times Q3\right) \\ & + \gamma_3 \left(Spillover_{sm} \times Q4\right) + \delta_1 Q3 + \delta_2 Q4 + \mu_b + \varepsilon_{psm}. \end{aligned} \quad (3)$$

We conduct two-tailed tests to reflect the theoretical uncertainty over the direction of the effects, although we expect to observe increases in incumbent party support in Q1/Q2 and decreases incumbent party support in Q4.

4.6 Ethical considerations

Our collaboration with the non-partisan NGO Borde Político followed prevailing ethical standards. First, we took great care to ensure that our evaluation of their campaign complied with all institutional and legal requirements for academic research and NGO activity. The study was approved by each of the three Institutional Review Boards at the universities of the authors, which includes a Mexican institution. Second, the intervention also complied with Mexican electoral law. As part of a similar Borde Político information campaign in 2015 (Arias, Larreguy, Marshall and Querubín 2019),[30] Mexican electoral authorities 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. We further corroborated this legal interpretation with a local electoral lawyer.

Beyond satisfying institutional and legal requirements, we regard our collaboration with Borde Político to evaluate their information campaign as both ethical and academically valuable for several reasons. First, the intervention evaluates the impact of information provision by 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. Moreover, all treatments were disseminated on behalf of Borde Político, who suggested delivering the information via Facebook to help understand how the effectiveness of their non-partisan campaigns could be maximized. The collaboration thus facilitated the first rigorous evaluation of the effect of campaign saturation within the context of the modern digital dissemination technologies that Borde Político—and many other NGOs—frequently use. The results may therefore be relevant to policymakers assessing the potential benefits and risks surrounding social media during election campaigns.

Second, all possible means were used to ensure that the campaign remained non-partisan. All numerical information, as well as the wording around it, was extracted from the independent and non-partisan ASF's online audit reports. In addition, the sample was chosen in a non-partisan manner: subject to a municipality holding elections in 2018 and the report pertaining to the current incumbent mayor's term in office, all municipalities for which a report was available entered our sample. To ensure that no party was differentially targeted, municipal treatment assignments were blocked on party. The ad itself further minimized perceptions of bias by avoiding the use of color

[30] Borde Político disseminated similar information through leaflets in that year.

schemes associated with any particular party.

Third, learning about the unstudied effects of high saturation campaigns inevitably requires concentrating information dissemination in ways that increase the possibility of affecting voting behavior. However, rather than providing citizens with fake or distracting information that could scramble their capacity to vote for their preferred candidate, the transparent and independent information provided by Borde Político was particularly relevant during an election when corruption was a salient issue for voters. Prior studies disseminating similar information show that Mexican voters care about the use of FISM funds (Arias, Balán, Larreguy, Marshall and Querubín 2019; Arias, Larreguy, Marshall and Querubín 2019; Chong et al. 2015; Larreguy, Marshall and Snyder 2019). Nevertheless, the ads—which did not mention the upcoming elections—did not explicitly ask citizens to respond in any way to the information. We thus expected that, to the extent that it reached voters, the information provided would help them to make better-informed voting decisions where the information was regarded as relevant.

Finally, we believe that it is important to inform Borde Político's commendable goal of improving municipal electoral accountability in a country where municipal governance is widely perceived to be tainted by corruption. Such goals often involve NGOs explicitly seeking to influence electoral outcomes. However, the chances of doing so in this particular context were slim because Facebook ads are often internalized by a small fraction of those targeted and because MORENA was widely expected to—and ultimately did—win by a landslide across much of the country.

5 Consumption and comprehension of Facebook ads

Before turning to the electoral results, we first examine the information campaign's reach, comprehension, and credibility. We show that Borde Político's Facebook ads achieved significant saturation in proportion with a municipality's intended saturation level, and that such information was internalized and generally regarded as credible by survey respondents.

5.1 The reach of the intervention

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 across treated municipalities. Table 1 presents Facebook's ad campaign-level analytics data by municipality, demonstrating that the campaign reached a considerable share of targeted adults in at least a limited way. Appendix Figure A3 plots trends in Facebook ad engagement by day, indicating that the campaign's reach increased over the course of the campaign's week.

Table 1 first shows that the ad campaign ultimately reached a substantial fraction of targeted

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 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: Each specification is estimated using OLS, and includes randomization block fixed effects. The two municipalities for which electoral data is unavailable are excluded. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

adults, broadly in proportion with the campaign's intended saturation level. Our estimates come from the following regression:

$$Y_m = \beta_1 Low\ saturation_m + \beta_2 High\ saturation_m + \mu_b + \varepsilon_m, \tag{4}$$

where $Low\ saturation_m$ and $High\ saturation_m$ are, respectively, indicators for the municipal-level low and high saturation treatment conditions. Column (1) reports that the Facebook ad appeared (on Facebook's "News Feed") as paid content 0.32 times for every adult member of the population in low saturation municipalities, and 0.95 times per adult in high saturation municipalities. Column (2) respectively reports a further 0.03 and 0.04 impressions per adult coming from organic views, which arose when friends on Facebook encountered the ad because Facebook ad viewers shared, commented, or reacted to (e.g. liked) an ad. To adjust for the number of citizens targeted, these numbers can be divided by the saturation level. Turning to unique Facebook users in columns (3), the campaign reached more than one third of its intended population in the average municipality, i.e. around half of the population of Facebook users. Column (4) again indicates that comparatively few additional views were generated organically. Although we cannot establish the intersection of respondents reached through paid-for ads and organic views or the location of organic views, these numbers imply that the mean Facebook user encountered the ad around three times. Appendix Figure A4 shows that, while there is variation in Facebook user engagement within assigned saturation levels, most 80% saturation municipalities experienced notably greater engagement than even the most-engaged 20% saturation municipalities.

A non-trivial number of Facebook users also engaged with the ad. Column (5) shows that 2-3% of targeted voting age adults clicked on the ad, e.g. by sharing, liking, or commenting on the ad or clicked through to the Facebook page. Furthermore, columns (7) and (9) respectively report that there were 0.04 views per adult of at least 3 seconds and 0.02 views per adult of at least 10 seconds in low saturation municipalities, while there were 0.11 and 0.06 such views per adult in high saturation municipalities. Column (11) shows that the corresponding numbers are 0.015 and 0.04 per adult, respectively, for the share of adults that watched the entire ad. The share of adults that watched for at least 10 seconds best approximates the share of citizens exposed to the information on the fraction of spending subject to irregularities, which was reported 17 seconds through the ad. Nevertheless, Facebook users that did not get that far could still have 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. Accounting for potential information diffusion and coordination that does not occur on Facebook, the ad campaign could have thus influenced a significant fraction of the electorate in treated segments and municipalities.

Although Facebook's algorithm ultimately dictates when ads appear to Facebook users, the intended 1:4 ratio of exposure to Borde Político's ads across low and high saturation treatments was generally maintained. The p values associated with the first test at the foot of panel A demonstrate that levels of engagement in high saturation municipalities were systematically greater by all non-organic metrics of engagement. While the high saturation treatment's reach was more limited relative to its target, 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 does not generally vary with the level of irregularities reported in the ad. Interacting the saturation variables in equation (4) with irregularities quartiles, panel B shows that—with the exception of high saturation municipalities with the highest amount of irregularities—citizen engagement did not 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 access to the Facebook ads. The exception in Q4 may reflect the relative lack of negative reactions from users on Facebook's interface to this information, which encourages Facebook's ad assignment algorithm to promote such ads. Our intent to treat estimates for the highest rates of irregularities thus imply lower effects of ad per view.

5.2 What types of Facebook users did the ads reach?

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 2 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 shows that 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.

The regression estimates in panels A and C further show that the Facebook ads reached similar

Table 2: Engagement with Facebook ads in treated municipalities, by demographic subgroup

	Reach 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. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

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 ad 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.

5.3 Comprehension of the treatment information

We next use the small-scale WhatsApp experiment conducted within our panel survey to examine whether respondents understood and internalized the ad's content a few weeks after receiving it. We leverage within-municipality variation in the receipt of a WhatsApp message containing the Facebook ad by estimating regressions of the form:

$$Y_{im} = \beta WhatsApp\ ad_{im} + \mu_m + \varepsilon_{im}, \quad (5)$$

where Y_{im} is an outcome for individual i in municipality m, and the municipality fixed effects, μ_m, mean that we exploit variation within the municipalities where the Facebook ads were delivered by WhatsApp message to survey respondents.[31] Robust standard errors are used in all analyses, and all observations are weighted by their inverse probability of treatment assignment.

Column (1) of Table 3 first shows that treated respondents were almost 9 percentage points more likely to report having received the WhatsApp message several weeks after the election ($p < 0.01$). Facebook analytics data further indicate that 32% of baseline respondents who received the WhatsApp message treatment clicked on the link to the Facebook page contained within the message.

Respondents also seem to have internalized the ad's content. Columns (2) and (3) report that treated respondents were 2.5 percentage points more likely to correctly identify (from a list of 4 options) that the message contained information about total FISM resources or the share subject to irregularities ($p < 0.05$) and were also 2 percentage points more likely to correctly identify (within 10 percentage point bands) the percentage of funds audited by the ASF that were subject to irregularities ($p < 0.01$). These estimates suggest that a quarter of the respondents who recalled the ad internalized the information. Since the post-election survey was conducted in the weeks after the election, this figure likely understates the share of respondents that recalled the information on election day. Moreover, panel B shows that the information was somewhat more memorable

[31] We do not examine the effects of municipal saturation due to differential attrition across municipality-level treatment conditions (see above).

Table 3: Comprehension and internalization of the information campaign ads

	Remembers WhatsApp message	**Knows content of message**	**Knows % of irregularities**
	(1)	(2)	(3)
Panel A: average treatment effects			
WhatsApp ad	0.089***	0.025**	0.024***
	(0.016)	(0.010)	(0.007)
R^2	0.083	0.051	0.061
Panel B: heterogeneity by irregularities quartile			
WhatsApp ad	0.044**	0.017	0.015*
	(0.020)	(0.013)	(0.009)
WhatsApp ad × Q3	0.109**	0.017	-0.002
	(0.052)	(0.032)	(0.022)
WhatsApp ad × Q4	0.119***	0.023	0.033**
	(0.036)	(0.023)	(0.015)
R^2	0.091	0.052	0.064
Observations	1,490	1,476	1,434
Outcome range	{0,1}	{0,1}	{0,1}
Control outcome mean	0.056	0.021	0.004
Control outcome std. dev.	0.229	0.144	0.059

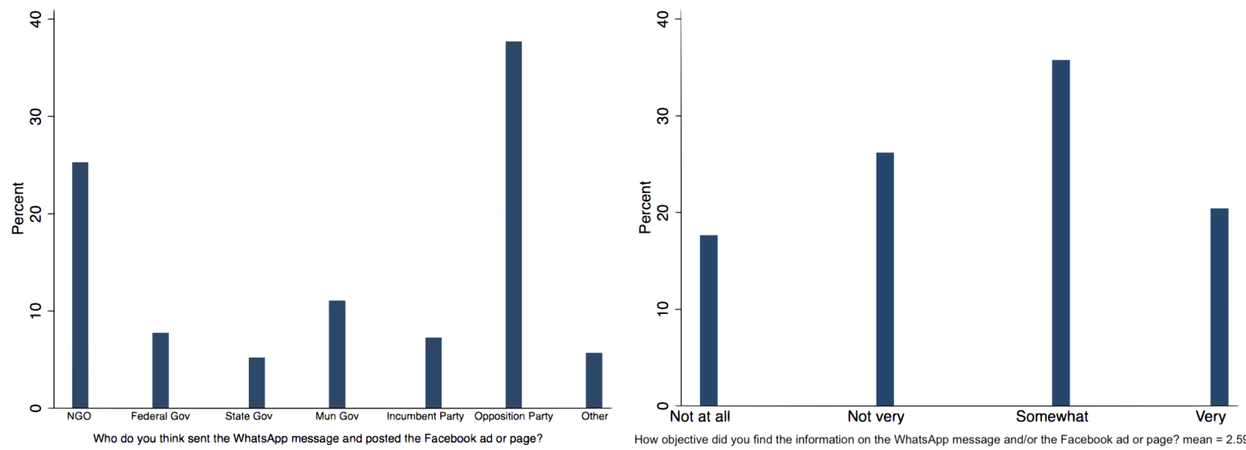
Notes: Each specification is estimated using OLS, and includes municipality fixed effects. All observations are weighted by the inverse probability of treatment assignment. Observations from municipalities where no WhatsApp treatment were distributed are excluded. Robust standard errors are in parentheses. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

when a large fraction of irregularities was reported, although this translated less systematically into accurate recall of the ad's content.

Together, these results suggest that the ad was likely to have been comprehensible to the many users who watched it on Facebook, and that it was likely to have been recalled by voters at the time of the election.

5.4 Perceived credibility of the ads

While the Facebook ads had significant reach and were likely to have been internalized by interested voters, they may have only influenced voter beliefs or behaviors if they were regarded as credible. We assess this by asking respondents that remembered receiving the WhatsApp message about its provenance and credibility. As Figure 4a illustrates, around a quarter of respondents cor-



(a) Perception of source

(b) Perception of credibility

Figure 4: Source and credibility of WhatsApp message/Facebook ad

Note: All data is from the subsample of endline respondents that remembered receiving the WhatsApp message.

rectly identified the ad's source as being from an NGO. Somewhat surprisingly given the generally favorable indicators of incumbent performance, one third regarded it as coming from opposition parties. Nevertheless, Figure 4b suggests that the majority of respondents regarded the information contained within the message as somewhat or very credible. Such levels of credibility are notable in an electoral context where many citizens were aware of and concerned by fake news. Moreover, the majority of respondents believed that such a Facebook ad campaign would be seen by many others in their municipality.

6 Effects of Facebook ads on voting behavior

We now present our precinct-level results. 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*. However, once we account for the share of irregularities reported, we find that the information campaign substantially increased the vote share of incumbent parties whose mayors were shown to have presided over zero or negligible irregularities. Furthermore, we demonstrate that these increases in support for the best-performing incumbents are greater among the treated segments within high saturation than within low saturation municipalities. We discuss spillover effects in the mechanisms section (see section 7).

Table 4: Effect of Facebook ads on precinct-level municipal incumbent party vote share (share of registered voters) and turnout

	Incumbent party vote (share of registered voters)		**Turnout**	
	(1)	(2)	(3)	(4)
Facebook ads	0.021*		0.013*	
	(0.011)		(0.008)	
Spillover	0.009		0.005	
	(0.011)		(0.008)	
Facebook ads in high saturation		0.024*		0.013
		(0.013)		(0.009)
Facebook ads in low saturation		0.007		0.001
		(0.013)		(0.010)
Spillover in high saturation		0.020		0.008
		(0.013)		(0.009)
Spillover in low saturation		0.005		0.002
		(0.013)		(0.009)
Observations	13,251	13,251	13,251	13,251
R^2	0.55	0.57	0.63	0.64
Control outcome mean	0.18	0.18	0.64	0.64
Control outcome std. dev.	0.11	0.11	0.12	0.12
Test: spillover $\geq$ direct (p value, 1-sided)	0.070		0.065	
Test within ads treatment: low $\geq$ high (p value, 1-sided)		0.110		0.081
Test within spillovers: low $\geq$ high (p value, 1-sided)		0.142		0.241

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

6.1 Average treatment effects

Table 4 presents estimates of the average treatment effect of the Facebook ad campaigns on municipal incumbent party vote share and turnout.[32] For each outcome, we examine the intent to treat effects of the segment-level treatment assignment. The block fixed effects and lagged dependent variables account for around 60% of the variation in our outcomes; this entails relatively precise standard errors of around a single percentage point.

We first find tentative evidence that the Facebook ads increased incumbent party support in the *average* municipality. Averaging across saturation and irregularities levels, column (1) shows that the Facebook ads treatment increased the incumbent party's vote share by 2.1 percentage points ($p < 0.1$) in treated segments. This 0.2 standard deviation increase represents around a 12% in-

[32]To save space, we restrict attention to incumbent party vote share, as a share of registered voters. Table A13 shows similar results for incumbent vote share, as a share of turnout.

crease in incumbent support, relative to the 18% of registered voters that turned out for the incumbent party in the control group. Column (2) reports a 2.4 percentage points effect in high saturation municipalities, although the difference relative to the 0.7 percentage point effect in low saturation municipalities is not quite statistically significant at the 10% level ($p = 0.11$, one-sided). These positive estimates are 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.[33] The next subsection shows that, as expected, the average effect pools across differential effects by the level of reported irregularities.

The Facebook ad campaign also slightly increased aggregate turnout. Column (3) reports a 1.3 percentage point increase in turnout due to the ad campaign ($p < 0.1$). Column (4) again indicates that this is predominantly driven by the Facebook ads in high saturation municipalities, as the differential between segments in high and low saturation municipalities is statistically significant ($p < 0.1$). The estimated effects on turnout are broadly in line with the positive effects of Facebook's own mobilization campaigns in the U.S. (see Bond et al. 2012; Jones et al. 2017). However, they remain small in magnitude, with the 0.1 standard deviation effect representing only a 2% increase in turnout relative to the baseline of 64% in the control group.

6.2 Heterogeneity by reported irregularities

The impact of providing Facebook ads reporting *any* level of incumbent expenditure irregularities resulted in a borderline statistically significant increase in support for the incumbent party. However, we next show that—as with many informational interventions—the relatively small average effects mask substantial heterogeneity with respect to the content of the information provided. We focus on the incumbent party vote, since the lack of differential effects on aggregate turnout (see Appendix Table A15) suggests that the results are driven primarily by shifts in municipal incumbent party support.

The results in Table 5 show that Borde Político's Facebook information campaign substantially increased the vote share of the least malfeasant municipal incumbent parties. In the municipalities where voting age adults in directly treated segments were informed of zero or negligible irregularities (i.e. quartiles Q1 and Q2, which are pooled because more than 25% of municipalities registered zero irregularities), column (1) shows that the Facebook ads treatment increased the incumbent party's vote share by 4.1 percentage points ($p < 0.05$)—an increase of more than 20%, relative to the control group mean. In contrast, the negative interaction coefficients for the third

[33] While performance indicators might often be expected to hurt incumbent parties that are already popular among voters (Ashworth, Bueno de Mesquita and Friedenberg 2018), the huge growth in support for MORENA between 2015 and 2018 may represent a sufficiently negative shock to the relative popularity of non-MORENA incumbents that the incumbent's electoral advantage might have been reversed. In our sample, there were only two MORENA incumbents.

Table 5: Effect of Facebook ads on precinct-level municipal incumbent party vote share (share of registered voters), by quartile of the sample irregularities distribution

	Incumbent party vote (share of registered voters)	
	(1)	(2)
Facebook ads	0.041**	0.057***
	(0.016)	(0.020)
Facebook ads × Q3	-0.046	-0.026
	(0.032)	(0.034)
Facebook ads × Q4	-0.051*	-0.079*
	(0.028)	(0.040)
Spillover	0.023	0.034*
	(0.015)	(0.018)
Spillover × Q3	-0.001	-0.008
	(0.026)	(0.031)
Spillover × Q4	-0.046	-0.071*
	(0.032)	(0.039)
Observations	13,251	13,251
R^2	0.59	0.65
Control outcome mean	0.18	0.18
Control outcome std. dev.	0.11	0.11
Test: null effect of Facebook ads in Q3 (p value, 2-sided)	0.849	0.244
Test: null effect of Facebook ads in Q4 (p value, 2-sided)	0.661	0.389
Test: null effect of spillover in Q3 (p value, 2-sided)	0.259	0.246
Test: null effect of spillover in Q4 (p value, 2-sided)	0.418	0.179
Interactive covariates		✓

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; amount of FISM funds received; 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/Q2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

and fourth quartiles (i.e. Q3 and Q4) indicate that the conditional average treatment effect is—as expected—smaller for higher levels of reported irregularities. The interaction coefficient capturing the difference in the effect of Facebook ads between Q1/Q2 and Q4 is statistically significant ($p < 0.1$). The tests at the foot of column (1) indicate that the overall effect in each quartile is only significantly different from zero for the municipal incumbent parties with the cleanest spending records.

The stronger electoral impact of good performance is in line with Arias, Larreguy, Marshall and Querubín's (2019) prior findings in a smaller set of Mexican municipalities in 2015. The limited sanctioning of higher levels of irregularities could reflect the counteracting benefit of reduced uncertainty about the incumbent's type among risk-averse voters (Arias, Larreguy, Marshall and Querubín 2019) or the possibility that bad performance is more likely to reach voters via the media. The limited sanctioning in Q4 could also reflect a lack of scope to reduce the incumbent party's vote share, given MORENA's electoral success in 2018 and the fact that MORENA was the incumbent in only one treated municipality. This is also consistent with Shadmehr and Bernhardt's (2017) prediction that the benefits of coordinating around signals of strong incumbent performance are greater where voters were already predisposed toward the challenger.

However, while access to ads is randomly assigned, the share of irregularities reported is not. To address the possibility that heterogeneity in response to Facebook ads across municipalities where the ASF found different levels of irregularities instead reflects other differences across these municipalities, we further adjust for the interaction between treatment conditions and 11 predetermined covariates at the municipal level. First, we include interactions with the other quantitative information conveyed by the ad—the financial year to which the audit pertained and the amount of FISM funds received by the municipality in that year—to address the possibility that other municipality-specific elements of the ad affected voting behavior. Second, we include the interaction with the prior municipal incumbent party vote share to address the possibility that audit report revelations differentially impact the parties of more or less popular local governments. Third, we further include interactions with eight demographic and socioeconomic variables that could both be correlated with reported irregularities and facilitate coordinated responses to ads or proxy for voters' prior beliefs, attentiveness or access to Facebook ads, or capacity to comprehend such ads.[34] Summary statistics for each covariate are shown in Appendix Table A3.

The estimates in column (2), which adjust for interactions between treatment conditions and all these potential confounds, are similar to the results in column (1) without such interactive covariates. Again, we find a large positive effect of Facebook ads in Q1/Q2 ($p < 0.01$). In Q4, we observe

[34]These municipal-level covariates are: the 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.

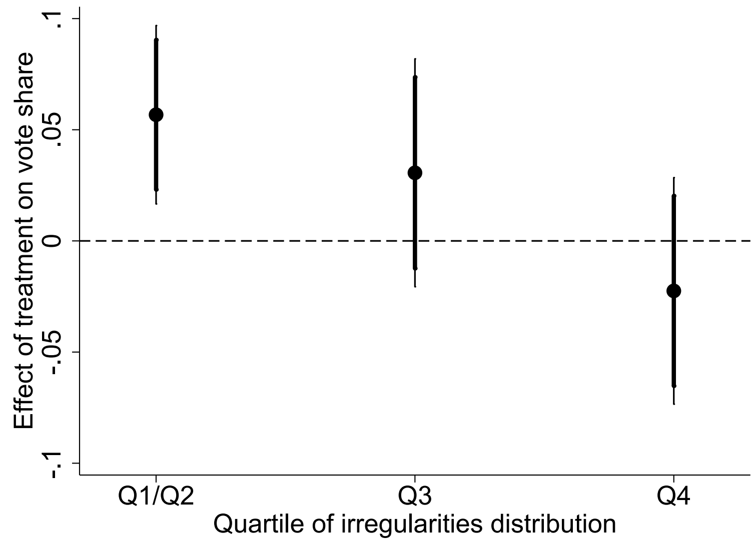


Figure 5: Conditional average treatment effect of Facebook ads, by irregularities quartile (with 90% and 95% confidence intervals)

Note: All estimates are from specification in column (2) of Table 5 with interactive covariates.

a larger decline in the incumbent party's vote share of 2.2 percentage points, although the test at the foot of the table indicates that this effect remains statistically insignificant at conventional levels.

The results are also robust to our decision to divide the irregularities distribution into quartiles. In particular, Appendix Table A16 reports similar increases in support for the best-performing incumbent parties when quartiles Q3 and Q4 are pooled to compare municipalities above and below the median level of irregularities or when irregularities are instead dichotomized as above 0% or above 5%. Moreover, when comparing municipalities above and below 5% irregularities, we also observe a statistically significant ($p < 0.1$) reduction in support for incumbent parties that oversaw irregularities that exceed 5% of audited expenditures.

The overall effects, which are illustrated graphically in Figure 5 for the irregularities by quartile specification that includes interactive covariates, are notable for their magnitude in comparison with previous findings. 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 costed approximately $11.62.[35] The 5.7 percentage point increase in incumbent party vote share in Q1/Q2 exceeds estimates from most prior experimental and quasi-experimental studies. Prior studies—which generally involve far lower levels of saturation—typically report effects of a couple of percentage points, even when murder charges

[35]Our back of the envelope calculation simply computes the average campaign spending per segment (the total campaign cost divided by the number of treated segments), and then divides this by the treatment effect in Q1/Q2: $\frac{\$17422.57}{263}\frac{1}{100*0.057} = \$11.62$. However, it is important to emphasize that the collection of ASF data, the production and targeting of the ads, and the credibility of Borde Político established through their prior work represent significant costs that the price of the ads purchased on Facebook does not reflect.

against candidates were reported via SMS in India (George, Gupta and Neggers 2019).[36] Only in the case of relatively widespread media reporting of similar types of malfeasance in Brazil (Ferraz and Finan 2008) have comparable impacts on vote shares been observed. As the next subsection shows, the large effects in our Mexican context are driven by the high saturation variant of the campaign.

6.3 Differential effects of information campaign saturation

Having established that voters targeted by Borde Político's accountability campaign rewarded "clean" municipal incumbent parties, we next examine the extent to which campaign saturation accentuates voters' electoral response to information provision. As argued above, large-scale information campaigns may stimulate belief updating and voter coordination that would not have occurred if incumbent performance information had only been provided to a small fraction of the electorate.

We test this hypothesis by leveraging the random assignment of treated municipalities to receive 20% or 80% information campaign saturation levels. Since Table 1 showed that the low saturation treatment was slightly more effective at reaching the citizens it targeted than the high saturation treatment, our estimates of equation (2) pertain to a case where effective saturation in high saturation municipalities was roughly 3 times greater than saturation in low saturation municipalities.

The results in columns (1) and (2) of Table 6 demonstrate that greater levels of ad saturation indeed amplify the effects of Facebook ads reporting zero or negligible irregularities. Regardless of whether interactions between treatment conditions and municipal-level covariates are adjusted for, reporting performance in Q1 or Q2 of the irregularities distribution increased the share of voters that voted for the incumbent party in treated segments by 5-6 percentage points ($p < 0.05$)—or around half a standard deviation—in high saturation municipalities. In contrast, the fourth row of the table shows that the effect in Q1/Q2 is slightly less than half this size in the directly treated segments within low saturation municipalities, and is statistically indistinguishable from zero. Furthermore, the tests at the foot of the table show that the 3 percentage point difference between the effects in directly treated segments within low and high saturation municipalities is statistically significantly ($p < 0.05$, one-tailed, when interactive covariates are adjusted for). Consistent with the lack of an effect in Q3 and Q4, we do not observe clear differential effects across saturation levels within Q3 and Q4.

These comparisons between directly treated segments and control municipalities indicate that a campaign's saturation can substantially increase its effect within areas directly targeted by Face-

[36]Most prior studies measure incumbent party vote share as a proportion of voters that turned out. The direct comparison in these cases is to Appendix Table A14, where normalizing by those that turned out increases the magnitude of the coefficients.

Table 6: Effect of Facebook ads on precinct-level municipal incumbent party vote share (share of registered voters), by information campaign saturation

	Incumbent party vote (share of registered voters)	
	(1)	(2)
Facebook ads in high saturation	0.049***	0.060**
	(0.018)	(0.026)
Facebook ads in high saturation × Q3	-0.069*	-0.044
	(0.038)	(0.041)
Facebook ads in high saturation × Q4	-0.062**	-0.049
	(0.031)	(0.054)
Facebook ads in low saturation	0.023	0.027
	(0.018)	(0.022)
Facebook ads in low saturation × Q3	-0.003	-0.032
	(0.029)	(0.036)
Facebook ads in low saturation × Q4	-0.048	-0.077*
	(0.035)	(0.046)
Spillover in high saturation	0.047**	0.056**
	(0.018)	(0.026)
Spillover in high saturation × Q3	-0.076*	-0.050
	(0.039)	(0.042)
Spillover in high saturation × Q4	-0.060**	-0.045
	(0.030)	(0.055)
Spillover in low saturation	0.014	0.020
	(0.017)	(0.021)
Spillover in low saturation × Q3	0.026	-0.002
	(0.029)	(0.035)
Spillover in low saturation × Q4	-0.039	-0.065
	(0.036)	(0.045)
Observations	13,251	13,251
R^2	0.63	0.70
Control outcome mean	0.18	0.18
Control outcome std. dev.	0.11	0.11
Test: larger effect of Facebook ads in high (vs. low) in Q1/Q2 (*p* value, 1-sided)	0.065	0.025
Test: larger effect of Facebook ads in high (vs. low) in Q3 (*p* value, 1-sided)	0.092	0.262
Test: larger effect of Facebook ads in high (vs. low) in Q4 (*p* value, 1-sided)	0.704	0.041
Test: larger effect of spillovers in high (vs. low) in Q1/Q2 (*p* value, 1-sided)	0.020	0.020
Test: larger effect of spillovers in high (vs. low) in Q3 (*p* value, 1-sided)	0.052	0.580
Test: larger effect of spillovers in high (vs. low) in Q4 (*p* value, 1-sided)	0.803	0.026
Interactive covariates		✓

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; amount of FISM funds received; 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/Q2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses.* denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests. The one-sided coefficient tests at the foot of the table test whether the effect of high saturation is larger in a given direction in high saturation than in low saturation municipalities.

book ads. This finding aligns with extant evidence suggesting that information campaigns are more likely to influence the voting behavior of treated voters when they are disseminated at a large scale by broadcast media (e.g. Ferraz and Finan 2008; Larreguy, Marshall and Snyder 2019), and is thus consistent with saturation driving the larger effects attributed to the mass media. While the framing and credibility functions of the media may also be important, our findings thus provide compelling evidence that saturation is central to the impact of information distributed via digital communication technologies. We next explore some of the potential mechanisms underpinning these effects.

7 Mechanisms

Our main results show that Facebook ads increased votes for the municipal incumbent parties that oversaw very low levels of expenditure irregularities, and that this effect was amplified in high saturation municipalities. If Borde Político's Facebook ad campaign only influenced those that viewed the ad directly on Facebook, our results suggest extremely high rates of persuasion. Indeed, the interpolated persuasion rate implied by assuming that only the voting behavior of the approximately 7% of targeted voters who reached the 17th second of the (paid-for or organically-generated) ad—when the share of irregularities appeared in the video—is 84%.[37] Even assuming that only 3 seconds of the ad is enough to increase support for the incumbent party by raising the issue's salience, the implied persuasion rate is still 39%. These rates far exceed those documented in prior studies of voting (DellaVigna and Gentzkow 2010), including the 3% persuasion rates associated with partisan ads on Facebook and Google in Germany (Rink 2019).

The relatively large effects of Borde Político's accountability campaign on voting behavior suggest either unusually high persuasion rates among atomized viewers of the ad, substantial interaction—such as through information diffusion or voter coordination—between voters that did and did not consume the ad, or that other actors capable of influencing large numbers of voters responded to Borde Político's Facebook ad campaign. We next document sizable spillovers within treated municipalities and limited responses to the information campaign from political campaigns and media outlets. Together, these findings suggest that social interactions between voters likely underpin the effectiveness of mass online information campaigns.

[37] Linear interpolation based on columns (9) and (11) of Table 1 suggests 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 that treated segments in high saturation municipalities are 4 times more prevalent than treated segments in low saturation municipalities in our sample, this yields a sample average exposure rate of around 6.9%. Following DellaVigna and Gentzkow (2010), the persuasion rate implied by column (2) of Table 5 is then: $100 \times \frac{0.057}{0.069}\frac{1}{1-0.18} \approx 84\%$.

7.1 Information campaigns and social interactions

Our survey data indicate that the information contained in Borde Político's campaign, and voters' reactions to it, could plausibly have influenced the vote choices of citizens that did not directly view the ad through social interactions. This would most likely operate through general political discussion between individuals. For instance, 76% of survey respondents reported discussing politics at least once a week with family and acquaintances in person or over the phone and 64% reported doing so through social media (i.e. Facebook, Twitter, and WhatsApp). Furthermore, 55% of survey respondents who recalled receiving the WhatsApp message reported discussing its content with their family and acquaintances. Suggesting that mass responses to the ads are possible, 66% of respondents suggested that discussions with others before the election helped them choose which candidate to vote for, while 84% reported that their expectations of others' vote choices affected their own vote choice. In contrast, the ad campaigns appear to have initiated relatively little activity online through Facebook's platform. As Table 1 shows, organic views of the ad only increased viewers beyond the paid-for Facebook ads by around 10%.

7.1.1 Estimates of spillover effects within treated municipalities

To examine whether the social interactions suggested by survey respondents drive the observed voting behavior, we compare segments within treated municipalities that were randomly assigned not to receive direct access to Facebook ads with segments within control municipalities. If Facebook's ad generation algorithm only reaches voters within targeted locations, voters in spillover segments could only have been influenced by Borde Político's Facebook ad campaign via interactions with voters in directly treated segments.

As shown in Table 5, the Facebook ads produced spillover effects that are only slightly smaller than the effects observed in directly-treated electoral precincts. Displaying the results graphically, Figure 6 shows that the presence of Facebook ads within the average treated municipality increased the incumbent party's vote share in untreated segments of municipalities in Q1/Q2 of the irregularities distribution by 3.4 percentage points ($p < 0.1$, when interactive covariates are adjusted for). This effect is around half the size of the effect in directly treated segments. Although we cannot observe how information was shared or what further discussions it facilitated, these estimates are consistent with a substantial effect of social interactions across directly and indirectly treated segments.

The spillover effect also increases with municipal saturation. Column (2) of Table 6 shows that the spillover effect rises to 5.6 percentage points in spillover segments within high saturation municipalities ($p < 0.05$), and falls to a statistically insignificant 2 percentage points in low saturation municipalities. As with the effects of Facebook ads in segments that were treated directly, the tests

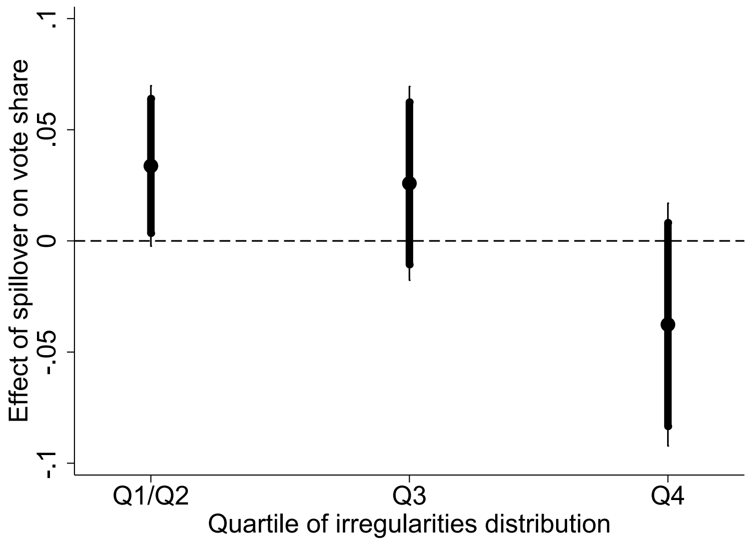


Figure 6: Conditional average treatment effect of spillovers, by irregularities quartile (with 90% and 95% confidence intervals)

Note: All estimates are from specification in column (2) of Table 5 with interactive covariates.

at the foot of Table 6 demonstrate that the greater spillover effect within high saturation municipalities, relative to low saturation municipalities, is also statistically significant ($p < 0.05$, one-sided test). This again suggests that explicit or tacit cross-segment interactions between citizens amplify an information campaign's effects. Both the direct effect of information *and* the associated spillovers thus increase in campaign saturation. The latter suggests that social interactions amplify effects throughout targeted municipalities, and thus also imply that social interactions could account for much of the effect within treated segments as well. If the difference between the effects in directly and indirectly treated segments is interpreted as the persuasion effect of watching the ad directly, our results suggest that most of the impact of Facebook ads is due to interactions occurring *after* an individual views the ad on Facebook.

7.1.2 Spillover effects are unlikely to reflect Facebook ad mistargeting

Most Facebook users saw Borde Político's informational ads on mobile devices providing GPS data that enable Facebook to target ads at users based on their home location. Nevertheless, spillover effects could still potentially reflect direct treatment arising from inaccurate geographic targeting of Facebook ads.[38] However, two tests suggest that mistargeting does not drive voting behavior.

First, if Facebook mistargeted ads to users that live—and thus vote—in untreated segments, we should expect to observe larger spillover effects in electoral precincts with access to 3G+ (3G, 4G, or LTE) mobile telecommunication signals. A lack of 3G technology renders videos on mobile

[38] Any such spillovers would, in contrast, underestimate the direct effects of Facebook ads.

devices slow and pixelated. However, this is likely to impede most other forms of social interaction far less. Restricting our sample to spillover and control segments, we assess this by examining the interaction between spillovers and an indicator for electoral precincts where 3G+ coverage reaches at least 75% of the population's home. The statistically insignificant differential effect estimates for the spillover effect in Q1/Q2 in columns (1) and (2) of Appendix Table A17 does not support the possibility of mistargeting.

Second, if treatments were mistargeted and voters considered politicians from the same party to be similarly malfeasant, we should expect to observe changes in the vote share of the incumbent party of a nearby treated municipality in precincts that are adjacent to that nearby treated municipality. To test for this possibility, while still leveraging experimental variation, we restrict our sample to precincts from non-experimental municipalities that are within 5 kilometers of only one experimental segment (see Appendix section A.13 for further details). Appendix Table A18 shows that proximity to neither (directly or indirectly) treated segments nor high or low municipal saturation levels in the nearby experimental municipality systematically affected the relevant vote share, within any irregularities quartile. These null estimates also suggest that Facebook ads were not significantly mistargeted.

7.2 Limited political and media responses to the Facebook ad campaigns

The main alternative class of mechanisms that could account for both the campaign's large direct and spillover effects and saturation's amplifying effects is the municipal-level strategies of other actors with the capacity to influence voters *en masse*. Most plausibly, the increase in incumbent party vote share in Q1/Q2 in directly and indirectly treated segments could result from incumbent parties or media outlets incorporating this information into large-scale campaign activities or news reports. Scope to respond to the intervention in this way was limited because the ads were distributed in the last week of the electoral campaign, while the results of the ASF reports were already available to parties and media outlets before Borde Político's intervention. Nevertheless, recent studies document sophisticated campaign responses to pre-election information dissemination campaigns (Arias, Larreguy, Marshall and Querubín 2019; Banerjee et al. 2011; Cruz, Keefer and Labonne 2019) and political debates (Bidwell, Casey and Glennerster forthcoming; Bowles and Larreguy 2019).

While we cannot determine exactly what occurred on the ground, we assess these alternative explanations by examining politician engagement with Borde Político's campaign online and reporting on the ASF reports (and corruption more generally) by local newspapers. We find little support for either alternative potential mechanism.

7.2.1 Online campaign responses to the Facebook ads

Online campaigning is now common in Mexico, where candidates use social media platforms to announce their campaign promises, publicize their slogans, and denounce other candidates. However, we were able to detect very few responses by mayoral candidates to the Facebook ads and associated pages.[39] Across all the Facebook and Twitter accounts that we identified as belonging to candidates in our sample of municipalities, we were only able to detect two responses to Borde Político's information campaign on Facebook and none on Twitter. The challenger candidate of the Alternative Sonora Movement in Huatabampo, Sonora, shared on his Facebook page that 30% of the FISM funds spent by the PRI municipal government were subject to irregularities. Similarly, the Citizen's Movement (MC) challenger candidate, eventually elected as municipal president in Venustiano Carranza, Michoacán, shared that the PRD municipal government incurred 14% in irregularities.

We additionally scraped thousands of comments, reactions, and shares relating to all Borde Político's Facebook ads and pages. Again, we observe negligible activity among Facebook users identified as running for other offices. A single PAN candidate for federal deputy liked the ad reporting 30% of irregularities by the PRI municipal government in Ciudad Valles, and one PRD candidate for federal deputy challenged the 61% of irregularities of the PRD municipal government reported in Cuautla, Morelos, arguing that there were no irregularities.[40] These scattered responses cannot plausibly account for the changes in voting behavior in treated segments at the scale that we observe.

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

To examine whether media reporting related to Borde Político's Facebook ad campaign amplified the campaign's direct effects and induced within-municipality spillovers, we collected online data from 263 local newspapers that serve 92 of the 124 municipalities in our final sample.[41] 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. While we could not obtain radio and television content, local newspapers are an important source of news content for local broadcast media outlets (see Larreguy, Lucas and Marshall 2016).

[39] Note that we cannot identify those Facebook users that shared and reacted to the Facebook ads and pages that do not have a public profile. However, this is unlikely to be a problem since candidates and political operatives are likely to have a public profile.

[40] Among the many other reactions by Facebook users, we were only able to identify 8 reactions that appeared to be from possible party operatives (people who regularly posted in favor of their party/candidate or explicitly mentioned working for the party or candidate's campaign).

[41] Since our randomization ensures that newspaper circulation within a given municipality is orthogonal to municipal treatment assignment and some small municipalities may not be served by local newspapers, we retain all municipalities for this analysis.

Over the 10 days between the start of the Facebook ad campaigns and the election, we searched for a variety of terms related to both the specific content of the Facebook ads and more general references to corruption. General references to corruption were included because the ad campaign could have increased demand for related information, which could also have influenced voting behavior if such demand was met by media outlets before the election. We estimate the effect of the campaign on such media reports using municipal-level regressions analogous to equation (4).

We again 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 in the pre-election period. Given that media outlets do sometimes report on the outcomes of ASF audits (Larreguy, Marshall and Snyder 2019), this suggests that the newspapers in the municipalities in our sample were unaware of the Facebook ads, had already reported on ASF (and did not see the need to discuss them further), or lacked incentives or resources to report on the issue during a nationwide election campaign. Second, we find that the ad campaigns did not significantly increase reporting more generally on issues related to corruption before the election. Indeed, Table 7 shows 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. These results suggest that the changes in vote choice induced by the Facebook ads were not driven by media coverage of the campaign.

7.3 Discussion of mechanisms behind changes in voting

We have shown that information campaign saturation most likely induced and amplified voter responses to Borde Político's Facebook ads through off-Facebook social interactions. Parsing whether social interactions changed voting behavior by inducing voters to update their beliefs about the incumbent's type or effort, or coordinate their vote choices, is challenging for several reasons that we now discuss. We also consider research designs that could distinguish between these mechanisms in future research.

First, the comparative statics implied by the belief updating and coordination mechanisms are often identical (Arias, Balán, Larreguy, Marshall and Querubín 2019; Zhuravskaya, Petrova and Enikolopov forthcoming). Any effects produced by either channel will generally increase in the extent of social interactions between citizens, e.g. through information diffusion or greater scope for coordination. While shutting down the coordination channel is essentially impossible in the field, ensuring that information about the incumbent is *fully internalized* by treated voters—such that further interaction with the information, due to informational spillovers, could not alter their beliefs about incumbent irregularities—is almost as difficult. Indeed, this cannot realistically be achieved at scale with mass media interventions that only provide citizens with access to informa-

Table 7: 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)
Any Saturation	-6.209	-5.483	-14.962				-6.209	-5.533	-15.042			
	(4.393)	(8.429)	(12.038)				(4.382)	(8.429)	(12.028)			
Any Saturation × Q3		-10.003	2.179					-9.803	2.352			
		(13.033)	(15.896)					(12.964)	(15.859)			
Any Saturation × Q4		4.608	25.325					4.593	25.391			
		(9.819)	(22.613)					(9.808)	(22.606)			
High saturation				-4.125	-4.548	-24.509				-4.173	-4.609	-24.494
				(4.746)	(9.965)	(17.266)				(4.735)	(9.965)	(17.256)
High saturation × Q3					-3.511	17.325					-3.390	17.381
					(13.155)	(23.292)					(13.121)	(23.260)
High saturation × Q4					0.998	54.470					0.875	54.246
					(12.537)	(33.992)					(12.504)	(33.982)
Low saturation				-8.370*	-6.636	-18.081				-8.321*	-6.676	-18.112
				(4.687)	(7.282)	(14.416)				(4.675)	(7.283)	(14.409)
Low saturation × Q3					-15.577	3.257					-15.305	3.449
					(15.274)	(19.237)					(15.174)	(19.190)
Low saturation × 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
R^2	0.37	0.40	0.60	0.37	0.42	0.68	0.37	0.40	0.60	0.37	0.42	0.68
Control outcome mean	8.51	8.51	8.51	8.51	8.51	8.51	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	31.45	31.45	31.45	31.45	31.45	31.45
Test: low = high (p value, 2-sided)				0.208						0.218		
Test: effect of treatment in Q3 (p value, 2-sided)		0.126	0.165					0.127	0.166			
Test: effect of treatment in Q4 (p value, 2-sided)		0.794	0.396					0.778	0.397			
Test: effect of low saturation in Q3 (p value, 2-sided)					0.123	0.222					0.125	0.225
Test: effect of low saturation in Q4 (p value, 2-sided)					0.710	0.251					0.704	0.251
Test: effect of high saturation in Q3 (p value, 2-sided)					0.216	0.585					0.215	0.587
Test: effect of high saturation in Q4 (p value, 2-sided)					0.608	0.109					0.587	0.111
Interactive covariates			✓			✓			✓			✓

Notes: Each specification is estimated using OLS, and includes randomization block fixed effects. The omitted irregularities category is Q1/Q2. Robust standard errors are in parentheses. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

tion.

However, as Chen and Yang (2019) show, it is possible to induce citizens to consume information. One approach to illuminating mechanisms other than belief updating about the incumbent's type or effort, induced by providing ASF audit results, would then be to incentivize a small number of voters to acquire and internalize the information provided by the mass campaign (e.g. Chen and Yang 2019), and then estimate treatment effects for compliers across municipalities with different levels of saturation or voters receiving different information about the campaign's scale. Alternatively, a detailed in-person explanation of the information may be even more effective at ensuring that some citizens could not plausibly update further after receiving the information from others (e.g. Bhandari, Larreguy and Marshall 2020; Boas, Hidalgo and Melo 2019; Humphreys and Weinstein 2012). Such designs could be implemented through parallel surveys, which could—under the (partially testable) assumption that voters do not update their beliefs on other margins after discussion with others—facilitate estimation of differential saturation effects that are not driven by information diffusion. Unfortunately, this study lacked the resources to implement such high-powered information engagement treatments.

Second, survey questions asking directly about particular mechanisms may struggle to distinguish between such mechanisms. The usual difficulties associated with conducting mediation analysis (see Bullock, Green and Ha 2010) are exacerbated in our context by the additional difficulty of ensuring that respondents engage with the mass-level version of the treatment that may be necessary for the activation of social interactions that occur outside the survey. Interpreting the survey evidence in our case was further limited by relatively low levels of engagement with the Facebook ads in our respondent sample and differential attrition across municipal treatment conditions. Nevertheless, we do observe some evidence consistent with both the belief updating and voter coordination: noisy estimates for belief updating from the WhatsApp treatment within our panel survey broadly align with voting behavior (see Appendix Table A19). Moreover, the descriptive data above suggests that information sharing and vote coordination are common in this setting.

8 Conclusions

We show that non-partisan information campaigns on social media can support electoral accountability, especially when the campaign's level of electorate saturation is high. Our evaluation of Borde Político's information campaign experimentally demonstrates that a large-scale Facebook ad campaign in Mexico substantially increased the vote share of municipal incumbent parties reported to have engaged in zero or negligible irregularities. This effect is particularly large among directly treated voters in high saturation municipalities, and appears to be driven by (off-Facebook) interactions between voters induced by the campaign, rather than responses by political campaigns or rebroadcasting by other media outlets.

These results suggest that the high saturation of many mass broadcast or social media messages may explain the greater impact of information dissemination conducted by mass media outlets. Appendix Table A1 more systematically documents this correlation across 15 studies. In this regard, our findings advance existing understandings of the effects of information dissemination on electoral accountability, on which extant evidence is decidedly mixed (Dunning et al. 2019). Of course, media outlets also 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 might more directly establish the extent to which saturation and these other potential mechanisms drive the role of mass media in promoting or hindering electoral accountability.

We show that information campaign saturation and social effects are key mechanisms underpinning information's effects on voting behavior. However, finer-grained distinctions between mechanisms, such as that between beliefs updating and tacit and explicit coordination, remain important topics for future research. Such distinctions may have important implications for policy makers seeking to establish in which contexts large-scale digital information provision could do most to increase accountability. Moreover, creative refinements to information campaigns could potentially be designed to accentuate particular interactions between citizens.

Finally, the relatively large effects of online information campaigns on vote shares pose interesting questions for democracy. While we focused on the potential benefits for electoral accountability of a non-partisan NGO campaign providing information gleaned from publicly-accessible independent audit reports, our findings also suggest that digital technologies could be used for more nefarious means. The use of social media around elections has recently come under intense scrutiny due to Russia's involvement in the 2016 U.S. presidential election, but also the frequent spreading of politically-relevant content via WhatsApp in the Global South. Our results suggest that such online information campaigns could have substantial electoral impacts, and thus highlight the importance of—and the possible trade-offs involved in—the decisions election regulators take to ensure that elections are not hijacked by fake news.

References

Acemoglu, Daron, Tarek A. Hassan and Ahmed Tahoun. 2017. "The Power of the Street: Evidence from Egypt's Arab Spring." *Review of Financial Studies* 31(1):1–42.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of The Nazis in Prewar Germany." *Quarterly Journal of Economics* 130(4):1885–1939.

Adida, Claire, Jessica Gottlieb, Eric Kramon and Gwyneth McClendon. 2019. "When Does Information Influence Voters? The Joint Importance of Salience and Coordination." *Comparative Political Studies, forthcoming* .

Alatas, Vivi, Abhijit Banerjee, Arun G Chandrasekhar, Rema Hanna and Benjamin A Olken. 2016. "Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia." *American Economic Review* 106(7):1663–1704.

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

Allcott, Hunt and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31(2):211–36.

Alt, James E., Amalie Jensen, Horacio Larreguy, David D. Lassen and John Marshall. 2020. "Diffusing Political Concerns: How Unemployment Information Passed Between Social Ties Influences Danish Voters." Working paper.

Anderson, Simon P. and John McLaren. 2012. "Media Mergers and Media Bias with Rational Consumers." *Journal of the European Economic Association* 10(4):831–859.

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. 2019. "Priors rule: When do Malfeasance Revelations Help or Hurt Incumbent Parties?" Working paper.

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.

Ashworth, Scott, Ethan Bueno de Mesquita and Amanda Friedenberg. 2018. "Learning about Voter Rationality." *American Journal of Political Science* 62(1):37–54.

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. 2019. "Public Information is an Incentive for Politicians: Experimental Evidence from Delhi Elections." Working paper.

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

Baron, David P. 2006. "Persistent media bias." *Journal of Public Economics* 90(1):1–36.

Barro, Robert J. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14(1):19–42.

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. 2020. "Able and mostly willing: An empirical anatomy of information's effect on voter-driven accountability in Senegal." Working paper.

Bidwell, Kelly, Katherine Casey and Rachel Glennerster. forthcoming. "Debates: Voting and Expenditure Responses to Political Communication." *Journal of Political Economy* .

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. 2019. "Who Debates, Wins? At-Scale Experimental Evidence on Debate Participation in a Liberian Election." Working paper.

Bullock, John G., Donald P. Green and Shang E. Ha. 2010. "Yes, But What's the Mechanism? (Don't Expect an Easy Answer)." *Journal of Personality and Social Psychology* 98(4):550–558.

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, Davide Cantoni, David Y. Yang, Noam Yuchtman and Y. Jane Zhang. 2019. "Persistent Political Engagement: Social Interactions and the Dynamics of Protest Movements." Working Paper.

Campante, Filipe, Ruben Durante and Francesco Sobbrio. 2017. "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation." *Journal of the European Economic Association* 16(4):1094–1136.

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

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.

Costas, Elena, Albert Solé-Ollé and Pilar Sorribas-Navarro. 2011. "Do voters really tolerate corruption? Evidence from Spanish Mayors." Working paper.

Cruz, Cesi, Philip Keefer and Julien Labonne. 2019. "Buying Informed Voters: New Effects of Information on Voters and Candidates." Working paper.

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

Da Silveira, Bernardo S. and Joao M.P. De Mello. 2011. "Campaign Advertising and Election Outcomes: Quasi-natural Experiment Evidence from Gubernatorial Elections in Brazil." *Review of Economic Studies* 78(2):590–612.

de Figueiredo, Miguel F.P., F. Daniel Hidalgo and Yuri Kasahara. 2014. "When Do Voters Punish Corrupt Politicians? Experimental Evidence from Brazil." Working paper.

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 Ethan Kaplan. 2007. "The Fox News effect: Media bias and voting." *Quarterly Journal of Economics* 122(3):1187–1234.

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

DellaVigna, Stefano, Ruben Enikolopov, Vera Mironova, Maria Petrova and Ekaterina Zhuravskaya. 2014. "Cross-border media and nationalism: Evidence from Serbian radio in Croatia." *American Economic Journal: Applied Economics* 6(3):103–32.

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, Maria Petrova and Aleksey Makarin. 2019. "Social Media and Protest Participation: Evidence from Russia." Econometrica, forthcoming.

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

Fafchamps, Marcel, Ana Vaz and Pedro C. Vicente. Forthcoming. "Voting and Peer Effects: Experimental Evidence from Mozambique." *Economic Development and Cultural Change* .

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.

Fergusson, Leopoldo and Carlos Molina. 2019. "Facebook Causes Protests." Working Paper.

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.

García-Jimeno, Camilo, Angel Iglesias and Pinar Yildirim. 2018. "Women, Rails and Telegraphs: An Empirical Study of Information Diffusion and Collective Action." NBER Working Paper 24495.

Gentzkow, Matthew, Jesse M. Shapiro and Michael Sinkinson. 2011. "The Effect of Newspaper Entry and Exit on Electoral Politics." *American Economic Review* 101(7):2980–3018.

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.

Gerber, Alan S., James G. Gimpel, Donald P. Green and Daron R. Shaw. 2011. "How Large and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment." *American Political Science Review* 105(1):135–150.

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.

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*

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

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.

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.

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. 2016. "When do Media Stations Support Political Accountability? A Field Experiment in Mexico." AEA RCT Registry AEARCTR-0001186.

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. 2018. "Leveling the playing field: How equalizing access to campaign advertising helps locally non-dominant parties in consolidating democracies." *Journal of the European Economic Association* 16(6):1812–1849.

Larreguy, Horacio, John Marshall and Jr. Snyder, James M. 2019. "Publicizing malfeasance: When the local media structure faciliates electoral accountability in Mexico." Working paper.

Larson, Jennifer M. 2017. "Networks and Interethnic Cooperation." *Journal of Politics* 79(2):546–559.

Levy, Ro'ee. 2019. "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." Working paper.

Liberini, Federica, Michela Redoanoz, Antonio Russo, Angel Cuevas and Ruben Cuevas. 2018. "Politics in the Facebook Era: Evidence from the 2016 US Presidential Elections." Working paper.

Lierl, Malte and Marcus Holmlund. forthcoming. Performance Information and Voting Behavior in Burkina Faso's Municipal Elections: Separating the Effects of Information Content and Information Delivery. In *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*, ed. Thad Dunning, Guy Grossman, Macartan Humphreys, Susan Hyde, Craig McIntosh and Gareth Nellis. Cambridge University Press.

Little, Andrew T. 2016. "Communication Technology and Protest." *Journal of Politics* 78(1):152–166.

Lynch, Marc. 2011. "After Egypt: The Limits and Promise of Online Challenges to the Authoritarian Arab State." *Perspectives on Politics* 9(2):301–310.

Manacorda, Marco and Andrea Tesei. forthcoming. "Liberation Technology: Mobile Phones and Political Mobilization in Africa." *Econometrica* .

Marshall, John. 2019*a*. "Signaling Sophistication: How Social Expectations Can Increase Political Information Acquisition." *Journal of Politics* 81(1):167–186.

Marshall, John. 2019*b*. "Tuning in, voting out: News consumption cycles, homicides, and electoral accountability in Mexico." Working paper.

Martin, Gregory J. and Ali Yurukoglu. 2017. "Bias in Cable News: Persuasion and Polarization." *American Economic Review* 107(9):2565–2599.

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

Morris, Stephen and Hyun Song Shin. 1998. "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks." *American Economic Review* pp. 587–597.

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

Nickerson, David W. 2008. "Is Voting Contagious? Evidence from Two Field Experiments." *American Political Science Review* 102(1):49–57.

Panagopoulos, Costas and Donald P. Green. 2008. "Field Experiments Testing the Impact of Radio Advertisements on Electoral Competition." *American Journal of Political Science* 52(1):156–168.

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

Pierskalla, Jan H. and Florian M. Hollenbach. 2013. "Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa." *American Political Science Review* 107(2):207–224.

Platas Izama, Melina and Pia Raffler. 2019. "Closing the Gap: Information and Mass Support in a Dominant Party Regime." Working paper.

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.

Rink, Anselm. 2019. "Do Online Ads Sway Elections? Evidence from a Field Experiment." Working Paper.

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.

Rogoff, Kenneth. 1990. "Equilibrium Political Budget Cycles." *American Economic Review* 80(1):21–36.

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

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

Spenkuch, Jorg L. and David Toniatti. 2018. "Political Advertising and Election Outcomes." *Quarterly Journal of Economics* 133(4):1981–2036.

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.

Vasudevan, Srinivasan. 2019. "Diminishing the Effectiveness of Vote Buying: Experimental Evidence from a Persuasive Radio Campaign in India." Working paper.

Zhuravskaya, Ekaterina, Maria Petrova and Ruben Enikolopov. forthcoming. "Political Effects of the Internet and Social Media." *Annual Review of Economics.*

A Appendix

Contents

A.1 Mixed evidence that information has improved electoral accountability

A substantial body of 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 the past decade.[42] 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.[43] 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 than exceed (fall below) prior expectations increase (decrease) support for the incumbent candidate or party;[44] 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.2 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 (2019) 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 (2019*b*) also finds that access to local media in Mexico drives sanctioning of municipal incumbents overseeing spikes in homicide rates before elections. Although 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

[42]For studies covering the Global North, see e.g. Berry and Howell (2007), Costas, Solé-Ollé and Sorribas-Navarro (2011), Fergusson (2014), Kendall, Nannicini and Trebbi (2015), and Snyder and Strömberg (2010).

[43]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 2019; Bidwell, Casey and Glennerster forthcoming; Platas Izama and Raffler 2019), on the basis that a wide range of information—which may or may not include incumbent performance indicators—is provided.

[44]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 (2019) 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 2019; Cruz et al. 2019) were excluded.

2002; Larreguy and Monteiro 2019; Reinikka and Svensson 2011) or the threat of future media reporting (Banerjee et al. 2019; Grossman and Michelitch 2018).

Although they also do not vary 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, Larreguy, Marshall and Querubín 2019), 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 2014).

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 forthcoming), or Uganda (Humphreys and Weinstein 2012). All but the last of these studies come from the recent Metaketa initiative, which coordinated similar accountability experiments across six developing countries and found negligible effects on average (Dunning et al. 2019).[45] Adida et al.'s (2019) 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 2020), 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-

[45]The Metaketa studies providing incumbent performance information are: Adida et al. (2019), Arias, Larreguy, Marshall and Querubín (2019), Boas, Hidalgo and Melo (2019), Buntaine et al. (2018), and Lierl and Holmlund (forthcoming); 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 (forthcoming).		
	Mixed findings	Adida et al. (2019).	Buntaine et al. (2018), Chong et al. (2015), de Figueiredo, Hidalgo and Kasahara (2014).	
	Mostly consistent findings	Bhandari, Larreguy and Marshall (2020).	Arias, Larreguy, Marshall and Querubín (2019), Banerjee et al. (2011), George, Gupta and Neggers (2019).	Ferraz and Finan (2008), Larreguy, Marshall and Snyder (2019), Marshall (2019*b*).

Notes: Underlined articles leverage experimental variation in access to information. See section A.2 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 that has experimentally varied electorate-level saturation introduced only limited variation in saturation and no experimental condition involved high levels of saturation.[46]

A.2 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 that 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, and discuss our coding decisions:

1. In Adida et al. (2019), we code the borderline statistically significant negative result for low dosage in column (2) of Table 4 that reports the impact of positive information on incumbent vote share 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 for high dosage in column (4) 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 had direct access to treatment. Overall, we thus code this study as providing "mixed findings."

2. In Arias, Larreguy, Marshall and Querubín (2019), we code the significant positive effect of the information treatment for 0% malfeasance and the significantly negative interaction between treatment and 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.

[46] Adida et al. (2019) compared 15 low dosage communes with 15 higher dosage communes, which reached 12-15% voters within around 4% and 30% of villages within each commune respectively.

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 (2020), 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 (2014), 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 (forthcoming), 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 (2019), and Marshall (2019*b*) 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.3 Infographics available on the Facebook 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.4 Additional information about WhatsApp treatments

Figure A2 shows the message received before the election by treated WhatsApp survey respondents. Treatments were sent via WhatsApp in the days before the election (after the baseline survey). All respondents in 20 of the 43 control municipalities received no WhatsApp treatments, while 80% of respondents in low- and high-saturation treated municipalities and the remaining 23 control municipalities received the WhatsApp treatment. To open the WhatsApp video, respondents needed to click to download the video. Within low- and high-saturation municipalities, these treatments were equally split between common knowledge and no common knowledge information, while all treated respondents in the 23 zero-saturation municipalities that received WhatsApp treatments did not receive the common knowledge treatment (since treatment had not been distributed at scale). These treatments were assigned within blocks of five similar respondents (based on baseline survey data) within municipalities.

A.5 Engagement with Facebook ads

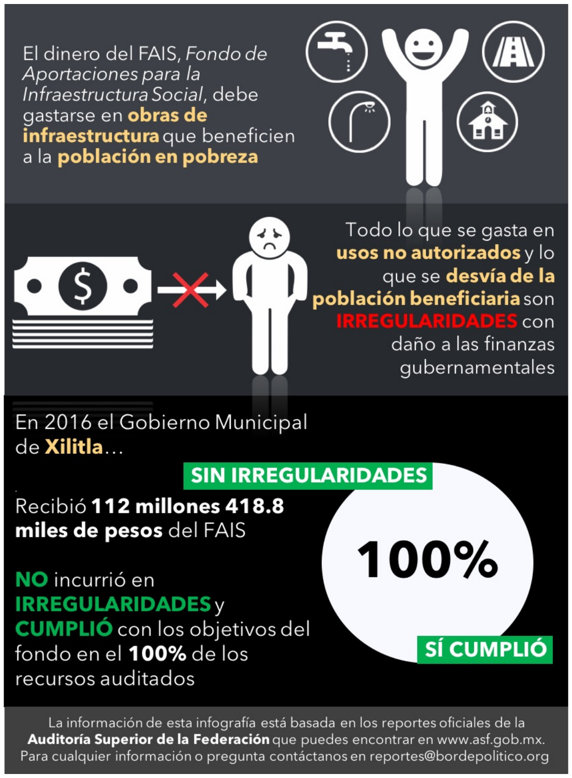
Figure A3 plots trends in Facebook user engagement with Borde Político's Facebook ad campaign over the pre-election period. We were not able to obtain day-by-day data on watching at least 10 seconds of the ad.

Figure A4 further reports the distribution of aggregate engagement with the Facebook ads within low and high saturation municipalities.

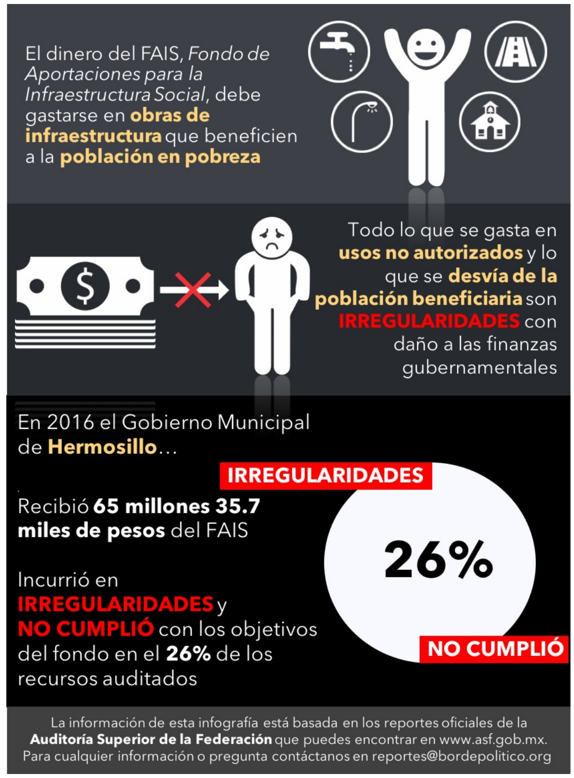
A.6 Deviation from the pre-analysis plan

All analyses follow our pre-analysis plan (available at socialscienceregistry.org/trials/3135), with the following exception:

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. In the presence of block fixed effects, the former issue could be



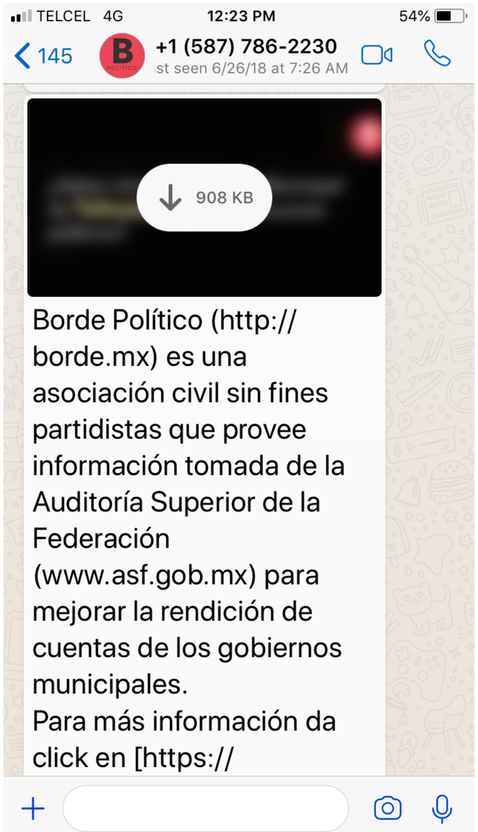
(a) Example of an infographic from a municipality (Xilitla, San Luis Potosí) where irregularities were 0%



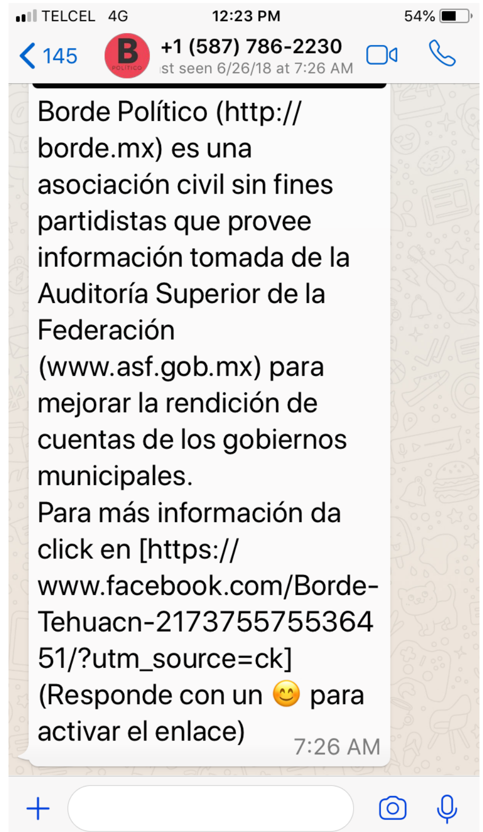
(b) Example of an infographic from a municipality (Hermosillo, Sonora) where irregularities were above 0%

Figure A1: Examples of treatment infographics

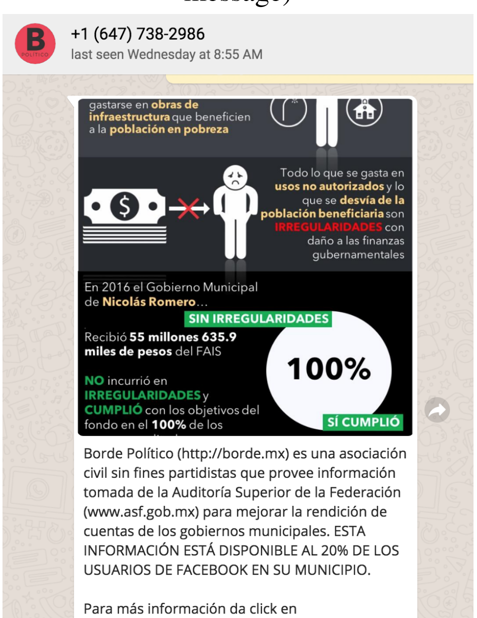
Note: In English, the first panel of the left infographic says "The funds from the FAIS, the Fund for Social Infrastructure, must be spent on infrastructure projects benefiting the poor;" the second panel says "All funds spent on unauthorized projects or not benefiting the poor constitute irregularities 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."



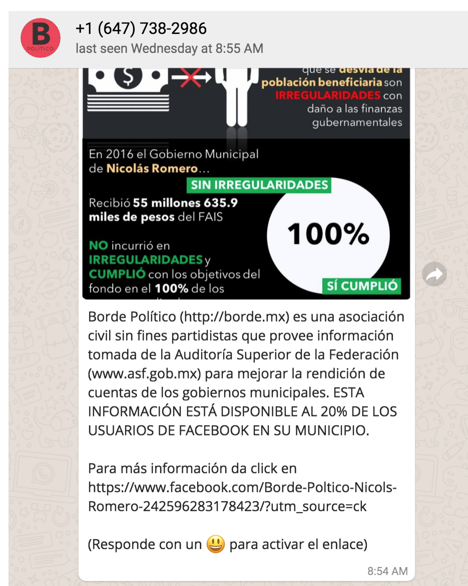
(a) WhatsApp message containing ad (top part of the message)



(b) WhatsApp message containing ad (bottom part of the message)

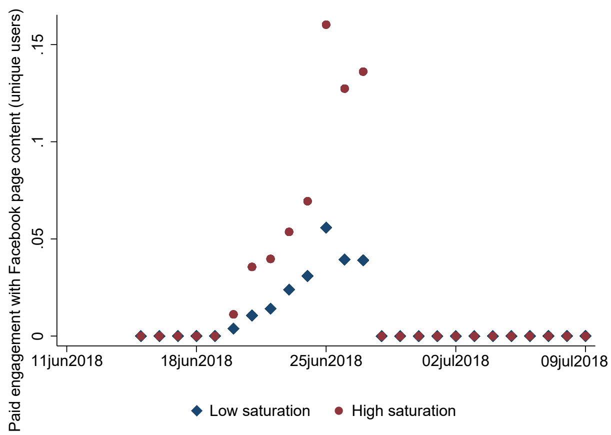


(c) Reminder WhatsApp message containing infographic (top)

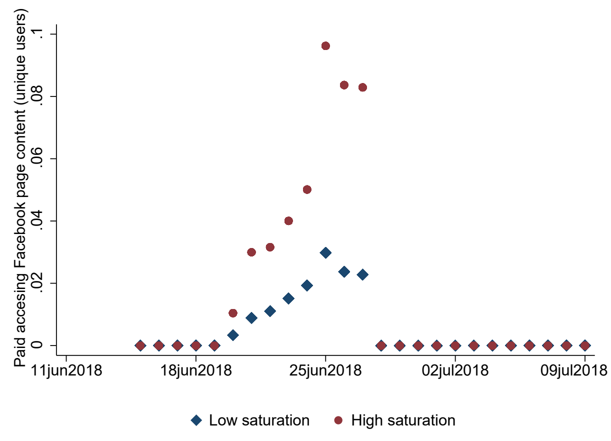


(d) Reminder WhatsApp message containing infographic (bottom)

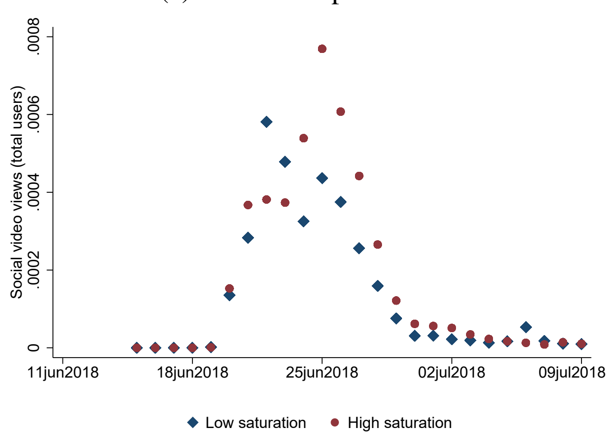
Figure A2: Example of the slides included in the ad video (from Nicolás Romero, Estado de México)



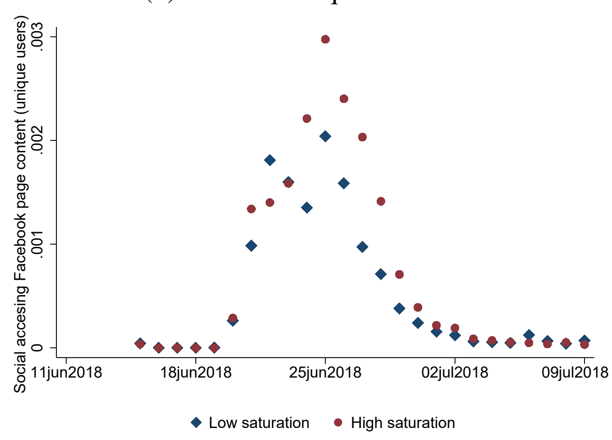
(a) Paid-for impressions



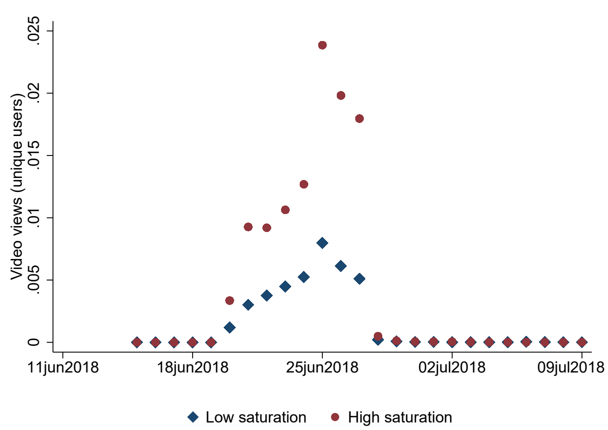
(b) Paid-for unique viewers



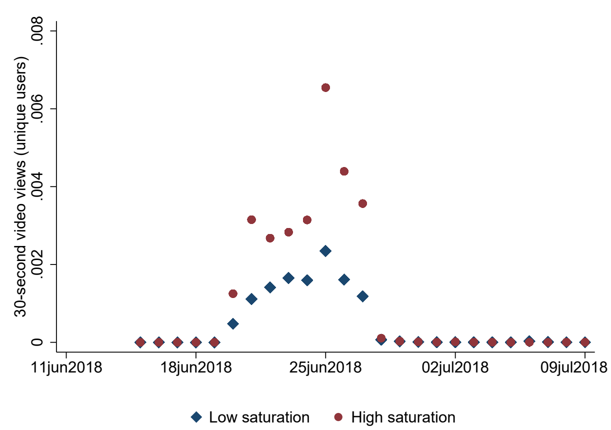
(c) Organic impressions



(d) Organic unique viewers

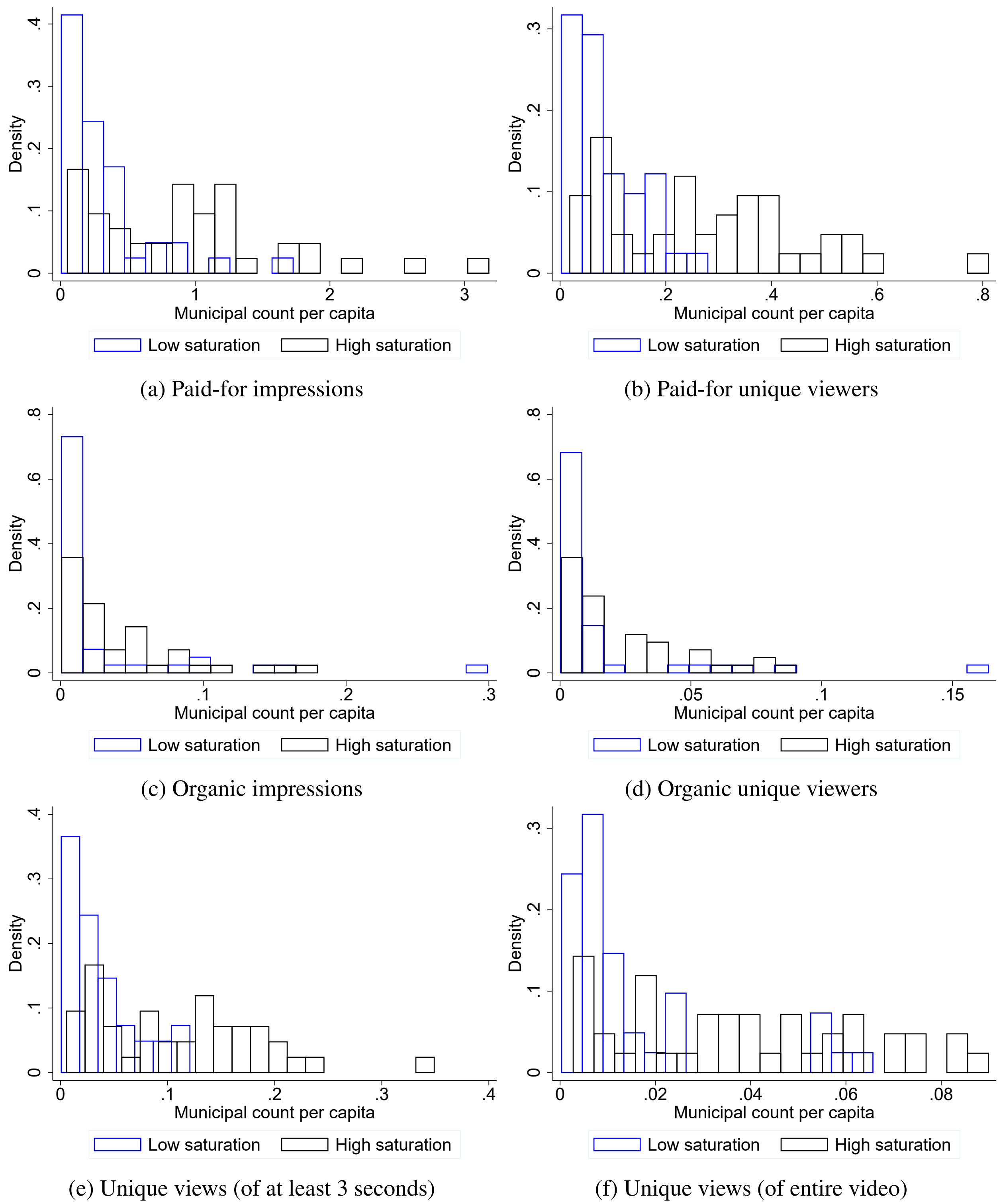


(e) Unique views (of at least 3 seconds)



(f) Unique views (of entire video)

Figure 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 addressed to yield unbiased estimates of ATEs and CATEs by interacting all treatments (and covariate interactions) with a fixed effect for this block. For simplicity, we choose not to do this for our main estimates; accordingly, no municipal-level weights are required. However, Table A2 reports similar results for our main estimates when these two municipalities are included in regressions that further use inverse weights to adjust for the different probability of treatment assignment in the 2-municipality block.

A.7 Balance tests

Table A3 shows balance tests based on the municipal-level treatment assignments in equation (4), where for each predetermined variable we report the p value associated with an F test of the restriction that $\beta_1 = \beta_2 = 0$. Only 3 of the 60 tests show a statistically significant difference at the $p < 0.1$ level.

Table A4 shows balance tests based on the segment-level treatment assignments in equation (2), where 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 1 of 62 tests show a statistically significant difference at the $p < 0.1$ level.

Tables A5 and A6 respectively report the municipal and individual level balance tests in our panel data. For the municipal-level Facebook ads treatment, Table A5 shows a number of statistically significant differences between the 0%, 20%, and 80% saturation conditions. Because of these differences in the types of respondent across municipalities, we do not leverage municipal-level variation in the Facebook ads treatment within our panel survey. As a consequence, we drop the municipalities where no WhatsApp treatments were administered to survey respondents and include municipality fixed effects in all analyses estimating the effect of the individual-level WhatsApp treatment. Excluding the municipal-level covariates that are perfectly explained by the municipality fixed effects, Table A6 reports balance across WhatsApp treatment groups in this subsample.

A.8 Alternative preregistered 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 Table 5, these decisions do not affect our findings. Indeed, the principal results in Table 5 are robust to excluding the lagged outcome (see Table A7), further weighting precincts by the number of registered voters (see Table A8), and weighting municipalities with different numbers of segments equally (see Table A9). The estimates without a lagged dependent variable (which reduce the R^2 by around 0.15) and

Table A2: Effect of Facebook ads on precinct-level municipal incumbent party vote share (share of registered voters), by quartile of the sample irregularities distribution and including the block containing only two municipalities

	Incumbent party vote (share of registered voters)			
	(1)	(2)	(3)	(4)
Facebook ads	0.040**	0.058***		
	(0.016)	(0.021)		
Facebook ads × Q3	-0.049	-0.031		
	(0.032)	(0.035)		
Facebook ads × Q4	-0.049*	-0.087**		
	(0.028)	(0.040)		
Spillover	0.023	0.035*		
	(0.015)	(0.018)		
Spillover × Q3	-0.005	-0.013		
	(0.026)	(0.031)		
Spillover × Q4	-0.045	-0.078**		
	(0.032)	(0.039)		
Facebook ads in high saturation			0.049***	0.065**
			(0.018)	(0.027)
Facebook ads in high saturation × Q3			-0.076**	-0.055
			(0.038)	(0.042)
Facebook ads in high saturation × Q4			-0.059*	-0.070
			(0.031)	(0.054)
Facebook ads in low saturation			0.024	0.031
			(0.018)	(0.022)
Facebook ads in low saturation × Q3			-0.010	-0.043
			(0.030)	(0.036)
Facebook ads in low saturation × Q4			-0.049	-0.090**
			(0.035)	(0.045)
Spillover in high saturation			0.044**	0.060**
			(0.018)	(0.027)
Spillover in high saturation × Q3			-0.081**	-0.060
			(0.039)	(0.043)
Spillover in high saturation × Q4			-0.055*	-0.065
			(0.030)	(0.055)
Spillover in low saturation			0.015	0.025
			(0.017)	(0.021)
Spillover in low saturation × Q3			0.019	-0.013
			(0.029)	(0.036)
Spillover in low saturation × Q4			-0.039	-0.078*
			(0.036)	(0.044)
Observations	13,274	13,274	13,274	13,274
R^2	0.59	0.64	0.62	0.70
Control outcome mean	0.18	0.18	0.18	0.18
Control outcome std. dev.	0.11	0.11	0.11	0.11
Test: null effect of Facebook ads in Q3 (*p* value, 2-sided)	0.710	0.297		
Test: null effect of Facebook ads in Q4 (*p* value, 2-sided)	0.693	0.265		
Test: null effect of spillover in Q3 (*p* value, 2-sided)	0.362	0.322		
Test: null effect of spillover in Q4 (*p* value, 2-sided)	0.426	0.122		
Test: larger effect of Facebook ads in high (vs. low) in Q1/Q2 (*p* value, 1-sided)			0.074	0.027
Test: larger effect of Facebook ads in high (vs. low) in Q3 (*p* value, 1-sided)			0.086	0.253
Test: larger effect of Facebook ads in high (vs. low) in Q4 (*p* value, 1-sided)			0.727	0.937
Test: larger effect of spillovers in high (vs. low) in Q1/Q2 (*p* value, 1-sided)			0.032	0.024
Test: larger effect of spillovers in high (vs. low) in Q3 (*p* value, 1-sided)			0.051	0.377
Test: larger effect of spillovers in high (vs. low) in Q4 (*p* value, 1-sided)			0.820	0.971
Interactive covariates		✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable,randomization block fixed effects, and an interaction between treatment. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; amount of FISM funds received; 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/Q2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses.* denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests. The one-sided coefficient tests at the foot of the table test whether the effect of high saturation is larger in a given direction in high saturation than in low saturation municipalities.

Table A3: 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	0.228	Share illiterate in 2010	0.13	0.11	0.440
FISM resources received	101,887.58	130,758.39	0.212	Share with no schooling in 2010	0.12	0.09	0.837
FISM resources used	90,442.80	114,967.96	0.158	Share with at least incomplete primary schooling in 2010	0.88	0.09	0.837
FISM resources audited	76,844.26	89,970.69	0.262	Share with at least complete primary schooling in 2010	0.71	0.16	0.649
Share of resources subject to irregularities	0.09	0.18	0.159	Share with at least incomplete secondary schooling in 2010	0.53	0.20	0.615
Municipal adult population in 2010	135,401.00	199,014.69	0.823	Share with at least complete secondary schooling in 2010	0.47	0.21	0.606
Registered voters in 2015	2,247.71	2,443.80	0.715	Share with higher education in 2010	0.25	0.19	0.502
Municipal incumbent vote share (reg. voters) in 2015	0.26	0.10	0.825	Share disabled in 2010	0.04	0.02	0.409
Municipal turnout in 2015	0.60	0.14	0.890	Share economically active in 2010	0.36	0.07	0.420
Federal MORENA-PT-PES vote share (reg. voters) in 2015	0.05	0.05	0.913	Share without health care in 2010	0.34	0.18	0.634
Federal PAN-PRD-MC vote share (reg. voters) in 2015	0.24	0.12	0.976	Share with state workers healthcare in 2010	0.04	0.05	0.117
Federal PRI-PVEM-NA vote share (reg. voters) in 2015	0.24	0.13	0.378	Average number of occupants per dwelling in 2010	4.27	0.63	0.474
Federal turnout in 2015	0.56	0.14	0.853	Average number of occupants per room in 2010	1.39	0.45	0.100*
Number of households in 2010	812.05	987.63	0.882	Share of households with more than 2 rooms in 2010	0.56	0.16	0.102
Private dwellings in 2010	1,030.64	1,416.93	0.761	Share of households with more than 3 rooms in 2010	0.65	0.19	0.101
Population in 2010	3,412.95	3,785.47	0.929	Share of households with non-dirt floor in 2010	0.89	0.12	0.464
Share women in 2010	0.51	0.02	0.293	Share of households with a toilet in 2010	0.89	0.17	0.032**
Share working age in 2010	0.61	0.06	0.491	Share of households with water in 2010	0.84	0.22	0.327
Share over 65 in 2010	0.06	0.03	0.791	Share of households with drainage in 2010	0.79	0.26	0.904
Share married in 2010	0.56	0.05	0.133	Share of households with electricity in 2010	0.95	0.08	0.418
Average children per woman in 2010	2.65	0.56	0.554	Share of households with water, drainage, and electricity in 2010	0.70	0.30	0.850
Share of households with a male head in 2010	0.78	0.08	0.166	Share of households with a washing machine in 2010	0.47	0.29	0.162
Share born out of state in 2010	0.12	0.15	0.553	Share of households with a landline telephone in 2010	0.29	0.24	0.319
Share Catholic in 2010	0.84	0.16	0.450	Share of households with a radio telephone in 2010	0.70	0.16	0.154
Share non-Catholic in 2010	0.11	0.13	0.147	Share of households with a fridge telephone in 2010	0.65	0.29	0.160
Share without religion in 2010	0.03	0.04	0.636	Share of households with a cell phone in 2010	0.47	0.30	0.585
Share indigenous speakers in 2010	0.21	0.35	0.059*	Share of households with a television in 2010	0.82	0.21	0.132
Average years of schooling in 2010	7.34	2.26	0.727	Share of households with a car in 2010	0.31	0.22	0.138
Average years of schooling for women in 2010	7.14	2.27	0.510	Share of households with a computer in 2010	0.18	0.19	0.558
Average years of schooling for men in 2010	7.55	2.28	0.782	Share of households with internet in 2010	0.12	0.16	0.560

Notes: The F tests report the p value associated with the test of the restriction that $\beta_1 = \beta_2 = 0$ in equation 4. Each specification is estimated using OLS, and includes randomization block fixed effects. All observations are weighted equally by municipality. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

Table A4: 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	0.396	Average years of schooling for men in 2010	7.90	2.42	0.886
FISM resources received	109,254.87	134,275.11	0.396	Share illiterate in 2010	0.12	0.11	0.185
FISM resources used	95,119.56	120,793.90	0.134	Share with no schooling in 2010	0.11	0.09	0.286
FISM resources audited	81,344.42	89,436.74	0.299	Share with at least incomplete primary schooling in 2010	0.89	0.09	0.286
Share of resources subject to irregularities	0.08	0.16	0.351	Share with at least complete primary schooling in 2010	0.73	0.16	0.448
Municipal adult population in 2010	209,177.31	274,260.97	0.875	Share with at least incomplete secondary schooling in 2010	0.56	0.21	0.767
Segment adult population in 2010	23,131.00	20,766.99	0.572	Share with at least complete secondary schooling in 2010	0.50	0.21	0.749
Precinct adult population in 2010	2,059.45	2,358.88	0.697	Share with higher education in 2010	0.28	0.21	0.849
Registered voters in 2015	2,314.20	2,594.90	0.796	Share disabled in 2010	0.04	0.02	0.639
Municipal incumbent vote share (reg. voters) in 2015	0.43	0.11	0.622	Share economically active in 2010	0.37	0.07	0.172
Municipal turnout in 2015	0.59	0.14	0.601	Share without health care in 2010	0.34	0.17	0.775
Federal MORENA-PT-PES vote share (reg. voters) in 2015	0.05	0.04	0.593	Share with state workers healthcare in 2010	0.04	0.05	0.274
Federal PAN-PRD-MC vote share (reg. voters) in 2015	0.23	0.11	0.935	Average number of occupants per dwelling in 2010	4.24	0.63	0.332
Federal PRI-PVEM-NA vote share (reg. voters) in 2015	0.23	0.12	0.364	Average number of occupants per room in 2010	1.33	0.45	0.101
Federal turnout in 2015	0.55	0.14	0.386	Share of households with more than 2 rooms in 2010	0.59	0.17	0.199
Number of households in 2010	827.99	1,036.54	0.800	Share of households with more than 3 rooms in 2010	0.68	0.20	0.132
Private dwellings in 2010	1,058.14	1,520.19	0.916	Share of households with non-dirt floor in 2010	0.90	0.12	0.426
Population in 2010	3,453.91	3,955.84	0.752	Share of households with a toilet in 2010	0.90	0.16	0.110
Share women in 2010	0.51	0.02	0.502	Share of households with water in 2010	0.85	0.22	0.135
Share working age in 2010	0.61	0.06	0.417	Share of households with drainage in 2010	0.82	0.25	0.193
Share over 65 in 2010	0.06	0.03	0.963	Share of households with electricity in 2010	0.96	0.08	0.234
Share married in 2010	0.55	0.05	0.761	Share of households with water, drainage, and electricity in 2010	0.73	0.30	0.372
Average children per woman in 2010	2.59	0.57	0.664	Share of households with a washing machine in 2010	0.51	0.30	0.120
Share of households with a male head in 2010	0.77	0.08	0.283	Share of households with a landline telephone in 2010	0.33	0.25	0.790
Share born out of state in 2010	0.12	0.15	0.323	Share of households with a radio telephone in 2010	0.72	0.16	0.238
Share Catholic in 2010	0.84	0.16	0.370	Share of households with a fridge telephone in 2010	0.68	0.29	0.133
Share non-Catholic in 2010	0.10	0.12	0.257	Share of households with a cell phone in 2010	0.51	0.30	0.187
Share without religion in 2010	0.03	0.04	0.994	Share of households with a television in 2010	0.84	0.20	0.126
Share indigenous speakers in 2010	0.18	0.33	0.050**	Share of households with a car in 2010	0.34	0.23	0.213
Average years of schooling in 2010	7.68	2.38	0.904	Share of households with a computer in 2010	0.21	0.21	0.898
Average years of schooling for women in 2010	7.47	2.38	0.786	Share of households with internet in 2010	0.15	0.18	0.919

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. Each specification is estimated using OLS, and includes randomization block fixed effects. All segments are weighted equally and by the inverse probability of treatment assignment. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

Table A5: Municipal-level Facebook treatment balance tests among endline survey respondents

	(1) Age	(2) Education	(3) Female	(4) Information acquisition index
High Saturation Municipal treatment	0.985	-0.105	-0.008	0.018
	(0.649)	(0.067)	(0.023)	(0.046)
Low Saturation Municipal treatment	0.595	-0.014	-0.028	0.032
	(0.713)	(0.060)	(0.026)	(0.039)
Observations	2,020	2,020	2,020	2,020
R^2	0.021	0.048	0.030	0.042
Control outcome mean	29.57	5.077	0.343	0.187
Control: Std.Dev.	12.17	0.932	0.475	0.572

	(5) Voted in last election	(6) Voted for incumbent	(7) Use of WhatsApp	(8) Use of Facebook
High Saturation Municipal treatment	0.041	0.038	-0.004*	-0.010*
	(0.033)	(0.042)	(0.002)	(0.005)
Low Saturation Municipal treatment	0.001	-0.054	-0.004*	-0.010*
	(0.033)	(0.042)	(0.002)	(0.006)
Observations	2,020	2,020	2,020	2,020
R^2	0.053	0.062	0.017	0.018
Control outcome mean	0.558	0.184	1	0.990
Control outcome std. dev.	0.468	0.387	0	0.098

	(9) Intend to vote	(10) Intend to vote for incumbent	(11) Others' intention to vote	(12) Others' intention to vote for incumbent
High Saturation Municipal treatment	3.470	1.791	-0.979	-2.339
	(2.476)	(4.126)	(1.178)	(2.007)
Low Saturation Municipal treatment	1.659	-1.772	-2.436*	-3.146*
	(2.088)	(3.586)	(1.396)	(1.664)
Observations	2,020	2,020	2,020	2,020
R^2	0.015	0.072	0.047	0.050
Control outcome mean	80.19	15.23	74.93	47.69
Control outcome std. dev.	39.89	33.40	17.36	26.12

	(13) Priors about irregularities	(14) Prior precision about irregularities	(15) Incumbent seeking reelection	(16) Amount of FISM funds audited
High Saturation Municipal treatment	-2.146	0.007	-0.049	-53,939.3***
	(1.564)	(0.020)	(0.165)	(20,380.4)
Low Saturation Municipal treatment	-0.961	0.019	0.109	-28,495.7
	(1.411)	(0.015)	(0.159)	(21,893.5)
Observations	2,020	2,020	2,020	2,020
R^2	0.025	0.054	0.441	0.645
Control outcome mean	62.94	0.044	0.385	81,749.3
Control outcome std. dev.	24.29	0.204	0.487	75,832.3

	(17) Adult Population	(18) Average years of schooling	(19) Share illiterate	(20) Average number of occupants per room
High Saturation Municipal treatment	-51,046.1	0.465**	-0.016**	-0.101**
	(52,867.0)	(0.232)	(0.008)	(0.048)
Low Saturation Municipal treatment	4,487.7	0.264	-0.015*	-0.111**
	(61,647.5)	(0.252)	(0.008)	(0.049)
Observations	2,020	2,020	2,020	2,020
R^2	0.813	0.692	0.622	0.588
Control outcome mean	399,423.1	8.983	0.058	1.114
Control outcome std. dev.	343,963.2	1.284	0.058	0.245

	(21) Average children per women	(22) Share of households with electricity, water and drainage	(23) Share of population that is working age	(24) Share of households with internet
High Saturation Municipal treatment	-0.117**	0.063**	0.010**	0.051**
	(0.055)	(0.029)	(0.004)	(0.023)
Low Saturation Municipal treatment	-0.089	0.036	0.011**	0.039
	(0.055)	(0.030)	(0.005)	(0.024)
Observations	2,020	2,020	2,020	2,020
R^2	0.600	0.609	0.688	0.708
Control outcome mean	2.275	0.852	0.644	0.238
Control outcome std. dev.	0.284	0.156	0.031	0.111

Notes: Each specification is estimated using OLS, and includes randomization block fixed effects. Standard errors clustered at the municipal level are in parentheses. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

Table A6: Individual-level WhatsApp treatment balance tests among endline survey respondents

	(1) Age	(2) Education	(3) Female	(4) Information acquisition index
WhatsApp ad	-0.264	-0.051	0.034	0.023
	(0.644)	(0.047)	(0.025)	(0.031)
Observations	1,490	1,490	1,490	1,490
R^2	0.071	0.092	0.053	0.082
Control outcome mean	29.65	5.036	0.329	0.153
Control outcome std. dev.	12.31	0.891	0.469	0.581

	(5) Voted in last election	(6) Voted for incumbent	(7) Use of WhatsApp	(8) Use of Facebook
WhatsApp ad	0.021	0.004	-0.003	-0.013**
	(0.024)	(0.018)	(0.002)	(0.005)
Observations	1,490	1,490	1,490	1,490
R^2	0.108	0.163	0.023	0.036
Control outcome mean	0.526	0.163	1	0.997
Control outcome std. dev.	0.461	0.370	0	0.059

	(9) Intend to vote	(10) Intend to vote for incumbent	(11) Others' intention to vote	(12) Others' intention to vote for incumbent
WhatsApp ad	4.365**	3.880**	0.281	2.341*
	(2.132)	(1.649)	(0.869)	(1.344)
Observations	1,490	1,490	1,490	1,490
R^2	0.066	0.130	0.105	0.110
Control outcome mean	77.08	12.43	74.37	45.39
Control outcome std. dev.	42.10	30.55	17.28	26.07

	(13) Priors about irregularities	(14) Prior precision about irregularities
WhatsApp ad	-0.141	0.001
	(1.357)	(0.012)
Observations	1,490	1,490
R^2	0.085	0.133
Control outcome mean	61.99	0.059
Control outcome std. dev.	24.94	0.236

Notes: Each specification is estimated using OLS, and includes municipality fixed effects. All observations are weighted by the inverse probability of treatment assignment. Observations from municipalities where no WhatsApp treatment were distributed are excluded. Robust standard errors are reported in parentheses. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

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

upweighting segments in smaller municipalities unsurprisingly yield slightly larger standard errors relative to our point estimates.

A.9 The common knowledge variant of the Facebook ads

As noted in the main text, some 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 A10 shows that the common knowledge treatment was no more likely to be engaged with than the non-common knowledge variant of the treatment. We observe no statistically significant difference in user interactions for the common knowledge campaigns. Second, Table A11 ultimately shows that no notable difference in the effect of the common knowledge and non-common knowledge 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.10 Null effects of municipal level treatments on incumbent election victory

Table A12 reports no discernible effect of the municipal-level treatments on whether the municipal incumbent party was re-elected.

Table A7: Effect of Facebook ads on precinct-level municipal incumbent party vote share (share of registered voters), by quartile of the sample irregularities distribution and excluding the lagged dependent variable

	Incumbent party vote (share of registered voters)			
	(1)	(2)	(3)	(4)
Facebook ads	0.036*	0.056***		
	(0.019)	(0.021)		
Facebook ads × Q3	-0.039	-0.027		
	(0.039)	(0.035)		
Facebook ads × Q4	-0.043	-0.061		
	(0.032)	(0.041)		
Spillover	0.014	0.033*		
	(0.018)	(0.019)		
Spillover × Q3	0.016	-0.014		
	(0.035)	(0.033)		
Spillover × Q4	-0.029	-0.052		
	(0.035)	(0.039)		
Facebook ads in high saturation			0.047**	0.055**
			(0.021)	(0.026)
Facebook ads in high saturation × Q3			-0.073*	-0.043
			(0.044)	(0.042)
Facebook ads in high saturation × Q4			-0.059	-0.020
			(0.037)	(0.056)
Facebook ads in low saturation			0.020	0.036
			(0.021)	(0.022)
Facebook ads in low saturation × Q3			0.022	-0.039
			(0.038)	(0.040)
Facebook ads in low saturation × Q4			-0.036	-0.059
			(0.039)	(0.047)
Spillover in high saturation			0.040*	0.048*
			(0.021)	(0.027)
Spillover in high saturation × Q3			-0.072	-0.045
			(0.048)	(0.043)
Spillover in high saturation × Q4			-0.053	-0.018
			(0.036)	(0.057)
Spillover in low saturation			0.005	0.020
			(0.021)	(0.021)
Spillover in low saturation × Q3			0.045	-0.013
			(0.037)	(0.038)
Spillover in low saturation × Q4			-0.018	-0.035
			(0.040)	(0.046)
Observations	13,278	13,278	13,278	13,278
R^2	0.45	0.57	0.51	0.62
Control outcome mean	0.18	0.18	0.18	0.18
Control outcome std. dev.	0.11	0.11	0.11	0.11
Test: null effect of Facebook ads in Q3 (p value, 2-sided)	0.919	0.295		
Test: null effect of Facebook ads in Q4 (p value, 2-sided)	0.800	0.834		
Test: null effect of spillover in Q3 (p value, 2-sided)	0.269	0.419		
Test: null effect of spillover in Q4 (p value, 2-sided)	0.616	0.499		
Test: larger effect of Facebook ads in high (vs. low) in Q1/Q2 (p value, 1-sided)			0.061	0.152
Test: larger effect of Facebook ads in high (vs. low) in Q3 (p value, 1-sided)			0.017	0.345
Test: larger effect of Facebook ads in high (vs. low) in Q4 (p value, 1-sided)			0.556	0.058
Test: larger effect of spillovers in high (vs. low) in Q1/Q2 (p value, 1-sided)			0.014	0.062
Test: larger effect of spillovers in high (vs. low) in Q3 (p value, 1-sided)			0.020	0.646
Test: larger effect of spillovers in high (vs. low) in Q4 (p value, 1-sided)			0.554	0.057
Interactive covariates		✓		✓

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; amount of FISM funds received; 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/Q2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses.* denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests. The one-sided coefficient tests at the foot of the table test whether the effect of high saturation is larger in a given direction in high saturation than in low saturation municipalities.

Table A8: Effect of Facebook ads on precinct-level municipal incumbent party vote share (share of registered voters), by quartile of the sample irregularities distribution and weighting by registered voters

	Incumbent party vote (share of registered voters)			
	(1)	(2)	(3)	(4)
Facebook ads	0.043**	0.060***		
	(0.018)	(0.021)		
Facebook ads × Q3	-0.058*	-0.044		
	(0.032)	(0.034)		
Facebook ads × Q4	-0.043*	-0.113**		
	(0.025)	(0.044)		
Spillover	0.025	0.041**		
	(0.017)	(0.020)		
Spillover × Q3	-0.026	-0.043		
	(0.028)	(0.033)		
Spillover × Q4	-0.033	-0.108**		
	(0.028)	(0.043)		
Facebook ads in high saturation			0.049**	0.069**
			(0.021)	(0.027)
Facebook ads in high saturation × Q3			-0.070*	-0.050
			(0.036)	(0.041)
Facebook ads in high saturation × Q4			-0.046*	-0.087
			(0.027)	(0.054)
Facebook ads in low saturation			0.019	0.031
			(0.020)	(0.023)
Facebook ads in low saturation × Q3			-0.035	-0.081**
			(0.033)	(0.036)
Facebook ads in low saturation × Q4			-0.034	-0.104**
			(0.030)	(0.047)
Spillover in high saturation			0.054**	0.073**
			(0.022)	(0.028)
Spillover in high saturation × Q3			-0.100**	-0.072*
			(0.040)	(0.042)
Spillover in high saturation × Q4			-0.051*	-0.079
			(0.028)	(0.058)
Spillover in low saturation			0.012	0.029
			(0.020)	(0.023)
Spillover in low saturation × Q3			0.003	-0.054
			(0.031)	(0.036)
Spillover in low saturation × Q4			-0.019	-0.095**
			(0.032)	(0.047)
Observations	13,251	13,251	13,251	13,251
R^2	0.60	0.66	0.64	0.74
Control outcome mean	0.17	0.17	0.17	0.17
Control outcome std. dev.	0.10	0.10	0.10	0.10
Test: null effect of Facebook ads in Q3 (*p* value, 2-sided)	0.561	0.523		
Test: null effect of Facebook ads in Q4 (*p* value, 2-sided)	0.998	0.085		
Test: null effect of spillover in Q3 (*p* value, 2-sided)	0.969	0.948		
Test: null effect of spillover in Q4 (*p* value, 2-sided)	0.688	0.036		
Test: larger effect of Facebook ads in high (vs. low) in Q1/Q2 (*p* value, 1-sided)			0.043	0.010
Test: larger effect of Facebook ads in high (vs. low) in Q3 (*p* value, 1-sided)			0.430	0.017
Test: larger effect of Facebook ads in high (vs. low) in Q4 (*p* value, 1-sided)			0.172	0.946
Test: larger effect of spillovers in high (vs. low) in Q1/Q2 (*p* value, 1-sided)			0.010	0.005
Test: larger effect of spillovers in high (vs. low) in Q3 (*p* value, 1-sided)			0.113	0.108
Test: larger effect of spillovers in high (vs. low) in Q4 (*p* value, 1-sided)			0.233	0.979
Interactive covariates		✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable, randomization block fixed effects, and an interaction between treatment. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; amount of FISM funds received; 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/Q2. All precincts are weighted voting age population and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses.* denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests. The one-sided coefficient tests at the foot of the table test whether the effect of high saturation is larger in a given direction in high saturation than in low saturation municipalities.

Table A9: Effect of Facebook ads on precinct-level municipal incumbent party vote share (share of registered voters), by quartile of the sample irregularities distribution and weighting municipalities with different numbers of segments equally

	Incumbent party vote (share of registered voters)			
	(1)	(2)	(3)	(4)
Facebook ads	0.037**	0.051**		
	(0.017)	(0.021)		
Facebook ads × Q3	-0.041	-0.013		
	(0.035)	(0.035)		
Facebook ads × Q4	-0.043	-0.062		
	(0.030)	(0.039)		
Spillover	0.020	0.024		
	(0.015)	(0.019)		
Spillover × Q3	0.004	0.008		
	(0.029)	(0.032)		
Spillover × Q4	-0.037	-0.052		
	(0.035)	(0.038)		
Facebook ads in high saturation			0.044**	0.053**
			(0.020)	(0.026)
Facebook ads in high saturation × Q3			-0.060	-0.028
			(0.041)	(0.041)
Facebook ads in high saturation × Q4			-0.060*	-0.034
			(0.033)	(0.052)
Facebook ads in low saturation			0.022	0.018
			(0.020)	(0.022)
Facebook ads in low saturation × Q3			0.003	-0.012
			(0.034)	(0.038)
Facebook ads in low saturation × Q4			-0.038	-0.059
			(0.039)	(0.046)
Spillover in high saturation			0.040**	0.047*
			(0.020)	(0.027)
Spillover in high saturation × Q3			-0.065	-0.031
			(0.042)	(0.042)
Spillover in high saturation × Q4			-0.057*	-0.027
			(0.032)	(0.053)
Spillover in low saturation			0.011	0.008
			(0.019)	(0.021)
Spillover in low saturation × Q3			0.034	0.021
			(0.034)	(0.038)
Spillover in low saturation × Q4			-0.029	-0.042
			(0.039)	(0.045)
Observations	13,251	13,251	13,251	13,251
R^2	0.58	0.65	0.62	0.70
Control outcome mean	0.18	0.18	0.18	0.18
Control outcome std. dev.	0.11	0.11	0.11	0.11
Test: null effect of Facebook ads in Q3 (p value, 2-sided)	0.899	0.155		
Test: null effect of Facebook ads in Q4 (p value, 2-sided)	0.800	0.662		
Test: null effect of spillover in Q3 (p value, 2-sided)	0.296	0.171		
Test: null effect of spillover in Q4 (p value, 2-sided)	0.586	0.314		
Test: larger effect of Facebook ads in high (vs. low) in Q1/Q2 (p value, 1-sided)			0.126	0.021
Test: larger effect of Facebook ads in high (vs. low) in Q3 (p value, 1-sided)			0.091	0.277
Test: larger effect of Facebook ads in high (vs. low) in Q4 (p value, 1-sided)			0.499	0.043
Test: larger effect of spillovers in high (vs. low) in Q1/Q2 (p value, 1-sided)			0.049	0.013
Test: larger effect of spillovers in high (vs. low) in Q3 (p value, 1-sided)			0.047	0.540
Test: larger effect of spillovers in high (vs. low) in Q4 (p value, 1-sided)			0.640	0.056
Interactive covariates		✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable, randomization block fixed effects, and an interaction between treatment. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; amount of FISM funds received; 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/Q2. All municipalities are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses.* denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests. The one-sided coefficient tests at the foot of the table test whether the effect of high saturation is larger in a given direction in high saturation than in low saturation municipalities.

Table A10: 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. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

Table A11: Differential effects of Facebooks ads in the 2018 municipal elections, by common knowledge treatment

	Incumbent party vote (share of turnout)			
	(1)	(2)	(3)	(4)
Facebook	0.020*	0.039**		
	(0.011)	(0.016)		
Facebook × Q3		-0.045		
		(0.033)		
Facebook × Q4		-0.042		
		(0.028)		
Facebook + CK	0.020*	0.042**		
	(0.011)	(0.016)		
Facebook + CK × Q3		-0.043		
		(0.031)		
Facebook + CK × Q4		-0.063**		
		(0.029)		
Spillover	0.008			
	(0.012)			
Spillover		0.022		
		(0.016)		
Spillover × Q3		0.004		
		(0.029)		
Spillover × Q4		-0.044		
		(0.032)		
Facebook ads in high saturation			0.024*	0.050**
			(0.013)	(0.019)
Facebook ads in high saturation × Q3				-0.066*
				(0.038)
Facebook ads in high saturation × Q4				-0.065**
				(0.032)
Facebook ads + CK in high saturation			0.023*	0.051***
			(0.013)	(0.019)
Facebook ads in low saturation			0.004	0.022
			(0.015)	(0.022)
Facebook ads in low saturation × Q3				-0.034
				(0.031)
Facebook ads in low saturation × Q4				-0.029
				(0.033)
Facebook ads + CK in low saturation			0.011	0.025
			(0.015)	(0.020)
Facebook ads + CK in low saturation × Q3				0.006
				(0.033)
Facebook ads + CK in low saturation × Q4				-0.072
				(0.043)
Spillover in high saturation			0.020	0.048**
			(0.013)	(0.019)
Spillover in low saturation			0.006	0.014
			(0.013)	(0.018)
Spillover in low saturation × Q3				0.028
				(0.031)
Spillover in low saturation × Q4				-0.039
				(0.036)
Spillover in high saturation × Q3				-0.077**
				(0.038)
Spillover in high saturation × Q4				-0.065**
				(0.031)
Facebook ads + CK in high saturation × Q3				-0.074**
				(0.036)
Facebook ads + CK in high saturation × Q4				-0.068**
				(0.032)
Observations	13,251	13,251	13,251	13,251
R^2	0.56	0.60	0.58	0.64
Control outcome mean	0.18	0.18	0.18	0.18
Control outcome std. dev.	0.11	0.11	0.11	0.11

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

Table A12: Effects of municipal saturation treatments on municipal election outcomes

	Incumbent party re-elected					
	(1)	(2)	(3)	(4)	(5)	(6)
Any Saturation	-0.078	-0.020				
	(0.076)	(0.133)				
Any Saturation × Q3		-0.065				
		(0.227)				
Any Saturation × Q4		-0.151				
		(0.232)				
High saturation			-0.047	0.062		
			(0.090)	(0.147)		
High saturation × Q3				-0.253		
				(0.285)		
High saturation × Q4				-0.253		
				(0.238)		
Low saturation			-0.109	-0.119		
			(0.092)	(0.157)		
Low saturation × Q3				0.152		
				(0.233)		
Low saturation × Q4				-0.042		
				(0.285)		
Saturation					-0.020	0.150
					(0.111)	(0.169)
Saturation × Q3						-0.412
						(0.333)
Saturation × Q4						-0.347
						(0.267)
Observations	124	124	124	124	124	124
R^2	0.48	0.53	0.48	0.55	0.47	0.54
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: low = high (p value, 2-sided)			0.534			
Test: effect of treatment in Q3 (p value, 2-sided)		0.587				0.331
Test: effect of treatment in Q4 (p value, 2-sided)		0.338				0.313
Test: effect of low saturation in Q3 (p value, 2-sided)				0.822		
Test: effect of low saturation in Q4 (p value, 2-sided)				0.474		
Test: effect of high saturation in Q3 (p value, 2-sided)				0.391		
Test: effect of high saturation in Q4 (p value, 2-sided)				0.276		

Notes: Each specification is estimated using OLS, and includes randomization block fixed effects. Robust standard errors are in parentheses. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

A.11 Results for other voting outcome measures

Table A13 reports the average treatment on incumbent vote share as a share of turnout.

Tables A14 and A15 report the treatment effects by irregularities quartile and saturation level for incumbent vote share, as a share of turnout, and turnout. The vote share results are similar to those reported in the main text, while the effects on turnout suggest that aggregate changes in turnout are not driving the aggregate changes in incumbent party vote share.

A.12 Robustness to alternative operationalizations of irregularities

Table A16 shows that the results are robust to using two alternative operationalizations of the content report. First, columns (1)-(4) compare municipalities with exactly 0% irregularities to municipalities with some irregularities. Second, columns (5)-(8) compare municipalities with at most 5% irregularities to municipalities with greater than 5% irregularities.

A.13 Tests of Facebook mistargeting

Tables A17 and A18 report the results of the geographic mistargeting tests described in the main text. Table A17 shows that precincts in spillover segments with high levels of 3G+ coverage are no more likely to change their voting behavior than precincts without good 3G+ coverage. As noted in the main article, this suggests that geographic mistargeting is unlikely to explain our findings because geographic mistargeting would most likely imply larger effects in nearby locations with easy access to Facebook.

Table A18 further examines cross-border spillovers to municipalities that should not have received Facebook ads. Specifically, we created a sample of precincts in non-experimental municipalities that are within 5 kilometers of an experimental (i.e. treated or control) segment. In order to avoid treatment correlating with the number of segments a cross-border spillover precinct is close to, we further restrict this sample to precincts for which only one experimental segment was within 5 kilometers. Consequently, we leverage the same experimental sources of variation used for the main analysis. This yielded a sample of 642 precincts within 5 kilometers of 184 different experimental segments from 88 of the experimental municipalities. Due to this smaller sample size, randomization block fixed effects are excluded (although the results are robust to their inclusion).

The results in columns (1) and (2) of Table A18 indicate that proximity of precincts in non-experimental municipalities to nearby segments that were (directly or indirectly) treated in experimental municipalities does not affect the precinct vote share of the incumbent party in the nearby experimental municipality. Columns (3) and (4) further show that the saturation level in the municipality of the nearby segment also does not affect the precinct vote share. Although this test

Table A13: Effect of Facebook ads on precinct-level municipal incumbent party vote share (share of turnout)

	Incumbent party vote (share of turnout)		
	(1)	(2)	(3)
High saturation	0.019		
	(0.018)		
Low saturation	0.015		
	(0.018)		
Facebook ads		0.026*	
		(0.016)	
Spillover		0.011	
		(0.016)	
Facebook ads in high saturation			0.031*
			(0.018)
Facebook ads in low saturation			0.010
			(0.018)
Spillover in high saturation			0.028
			(0.019)
Spillover in low saturation			0.008
			(0.019)
Observations	13,251	13,251	13,251
R^2	0.50	0.51	0.53
Control outcome mean	0.28	0.28	0.28
Control outcome std. dev.	0.15	0.14	0.14
Test: low $\geq$ high (p value, 1-sided)	0.832		
Test: spillover $\geq$ direct (p value, 1-sided)		0.107	
Test within ads treatment: low $\geq$ high (p value, 1-sided)			0.143
Test within spillovers: low $\geq$ high (p value, 1-sided)			0.158

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable and randomization block fixed effects. All segments are weighted equally and by the inverse probability of (municipal or segment, as appropriate) treatment assignment. Standard errors clustered by municipality are in parentheses. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.

Table A14: Effect of Facebook ads on precinct-level municipal incumbent party vote share (share of turnout), by quartile of the sample irregularities distribution

	Incumbent party vote (share of turnout)			
	(1)	(2)	(3)	(4)
Facebook ads	0.058	0.088		
	(0.022)	(0.028)		
Facebook ads × Q3	-0.087	-0.071		
	(0.044)	(0.047)		
Facebook ads × Q4	-0.061	-0.118		
	(0.038)	(0.055)		
Spillover	0.029	0.055		
	(0.021)	(0.026)		
Spillover × Q3	0.000	-0.027		
	(0.039)	(0.045)		
Spillover × Q4	-0.047	-0.096		
	(0.043)	(0.054)		
Facebook ads in high saturation			0.074	0.092
			(0.026)	(0.035)
Facebook ads in high saturation × Q3			-0.136	-0.097
			(0.053)	(0.058)
Facebook ads in high saturation × Q4			-0.079	-0.078
			(0.044)	(0.073)
Facebook ads in low saturation			0.025	0.046
			(0.026)	(0.029)
Facebook ads in low saturation × Q3			0.003	-0.072
			(0.044)	(0.052)
Facebook ads in low saturation × Q4			-0.038	-0.086
			(0.050)	(0.065)
Spillover in high saturation			0.072	0.086
			(0.026)	(0.036)
Spillover in high saturation × Q3			-0.142	-0.096
			(0.054)	(0.059)
Spillover in high saturation × Q4			-0.079	-0.069
			(0.043)	(0.075)
Spillover in low saturation			0.012	0.034
			(0.025)	(0.028)
Spillover in low saturation × Q3			0.048	-0.017
			(0.044)	(0.052)
Spillover in low saturation × Q4			-0.029	-0.074
			(0.050)	(0.063)
Observations	13,251	13,251	13,251	13,251
R^2	0.56	0.63	0.59	0.68
Control outcome mean	0.28	0.28	0.28	0.28
Control outcome std. dev.	0.14	0.14	0.14	0.14
Test: null effect of Facebook ads in Q3 (*p* value, 2-sided)	0.412	0.625		
Test: null effect of Facebook ads in Q4 (*p* value, 2-sided)	0.916	0.405		
Test: null effect of spillover in Q3 (*p* value, 2-sided)	0.352	0.404		
Test: null effect of spillover in Q4 (*p* value, 2-sided)	0.626	0.294		
Test: larger effect of Facebook ads in high (vs. low) in Q1/Q2 (*p* value, 1-sided)			0.030	0.026
Test: larger effect of Facebook ads in high (vs. low) in Q3 (*p* value, 1-sided)			0.016	0.677
Test: larger effect of Facebook ads in high (vs. low) in Q4 (*p* value, 1-sided)			0.586	0.143
Test: larger effect of spillovers in high (vs. low) in Q1/Q2 (*p* value, 1-sided)			0.010	0.017
Test: larger effect of spillovers in high (vs. low) in Q3 (*p* value, 1-sided)			0.010	0.448
Test: larger effect of spillovers in high (vs. low) in Q4 (*p* value, 1-sided)			0.688	0.055
Interactive covariates		✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable, randomization block fixed effects, and an interaction between treatment. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; amount of FISM funds received; 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/Q2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses.* denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests. The one-sided coefficient tests at the foot of the table test whether the effect of high saturation is larger in a given direction in high saturation than in low saturation municipalities.

Table A15: Effect of Facebook ads on precinct-level municipal turnout, by quartile of the sample irregularities distribution

	Turnout			
	(1)	(2)	(3)	(4)
Facebook ads	0.008	-0.007		
	(0.012)	(0.011)		
Facebook ads × Q3	0.025	0.046		
	(0.020)	(0.017)		
Facebook ads × Q4	-0.012	0.023		
	(0.022)	(0.023)		
Spillover	0.006	-0.008		
	(0.013)	(0.010)		
Spillover × Q3	0.005	0.020		
	(0.020)	(0.018)		
Spillover × Q4	-0.019	0.001		
	(0.020)	(0.020)		
Facebook ads in high saturation			0.004	-0.002
			(0.013)	(0.011)
Facebook ads in high saturation × Q3			0.038	0.043
			(0.023)	(0.019)
Facebook ads in high saturation × Q4			-0.006	0.030
			(0.025)	(0.025)
Facebook ads in low saturation			0.006	-0.010
			(0.014)	(0.011)
Facebook ads in low saturation × Q3			0.004	0.036
			(0.023)	(0.018)
Facebook ads in low saturation × Q4			-0.027	-0.009
			(0.024)	(0.021)
Spillover in high saturation			-0.001	-0.004
			(0.013)	(0.012)
Spillover in high saturation × Q3			0.032	0.031
			(0.023)	(0.020)
Spillover in high saturation × Q4			-0.001	0.026
			(0.023)	(0.025)
Spillover in low saturation			0.007	-0.005
			(0.014)	(0.011)
Spillover in low saturation × Q3			-0.003	0.014
			(0.023)	(0.019)
Spillover in low saturation × Q4			-0.022	-0.011
			(0.023)	(0.022)
Observations	13,251	13,251	13,251	13,251
R^2	0.63	0.69	0.65	0.71
Control outcome mean	0.64	0.64	0.64	0.64
Control outcome std. dev.	0.12	0.12	0.12	0.12
Test: null effect of Facebook ads in Q3 (p value, 2-sided)	0.026	0.001		
Test: null effect of Facebook ads in Q4 (p value, 2-sided)	0.792	0.364		
Test: null effect of spillover in Q3 (p value, 2-sided)	0.430	0.343		
Test: null effect of spillover in Q4 (p value, 2-sided)	0.383	0.628		
Test: larger effect of Facebook ads in high (vs. low) in Q1/Q2 (p value, 1-sided)			0.582	0.863
Test: larger effect of Facebook ads in high (vs. low) in Q3 (p value, 1-sided)			0.024	0.120
Test: larger effect of Facebook ads in high (vs. low) in Q4 (p value, 1-sided)			0.782	0.001
Test: larger effect of spillovers in high (vs. low) in Q1/Q2 (p value, 1-sided)			0.223	0.549
Test: larger effect of spillovers in high (vs. low) in Q3 (p value, 1-sided)			0.053	0.241
Test: larger effect of spillovers in high (vs. low) in Q4 (p value, 1-sided)			0.782	0.045
Interactive covariates		✓		✓

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable, randomization block fixed effects, and an interaction between treatment. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; amount of FISM funds received; 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/Q2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses.* denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests. The one-sided coefficient tests at the foot of the table test whether the effect of high saturation is larger in a given direction in high saturation than in low saturation municipalities.

Table A16: Effect of Facebook ads on precinct-level municipal incumbent party vote share (share of registered voters), by binary operationalizations of irregularities

	Incumbent party vote (share of registered voters)											
	Higher malfeasance measure:											
	Pooling Q3 and Q4				>0% irregularities				>5% irregularities			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Facebook ads	0.044***	0.055***			0.042**	0.053**			0.033**	0.051***		
	(0.017)	(0.020)			(0.017)	(0.021)			(0.014)	(0.016)		
Facebook ads × Higher malfeasance	-0.057**	-0.060*			-0.048*	-0.051			-0.068**	-0.094***		
	(0.025)	(0.033)			(0.025)	(0.032)			(0.027)	(0.034)		
Spillover	0.022	0.030			0.023	0.029			0.024*	0.032**		
	(0.015)	(0.018)			(0.015)	(0.018)			(0.012)	(0.014)		
Spillover × Higher malfeasance	-0.029	-0.046			-0.028	-0.041			-0.071**	-0.090***		
	(0.025)	(0.031)			(0.025)	(0.031)			(0.030)	(0.033)		
Facebook ads in high saturation			0.054***	0.064**			0.050**	0.057**			0.037**	0.049**
			(0.019)	(0.025)			(0.019)	(0.025)			(0.016)	(0.020)
Facebook ads in high saturation × Higher malfeasance			-0.076***	-0.057			-0.061**	-0.037			-0.079***	-0.054
			(0.029)	(0.041)			(0.029)	(0.041)			(0.029)	(0.044)
Facebook ads in low saturation			0.021	0.026			0.024	0.024			0.026*	0.018
			(0.019)	(0.020)			(0.019)	(0.020)			(0.015)	(0.016)
Facebook ads in low saturation × Higher malfeasance			-0.030	-0.062*			-0.035	-0.052			-0.076**	-0.084**
			(0.028)	(0.034)			(0.028)	(0.034)			(0.034)	(0.038)
Spillover in high saturation			0.051***	0.060**			0.048**	0.055**			0.033**	0.043**
			(0.019)	(0.025)			(0.019)	(0.026)			(0.016)	(0.020)
Spillover in high saturation × Higher malfeasance			-0.079***	-0.059			-0.067**	-0.044			-0.076**	-0.046
			(0.029)	(0.042)			(0.028)	(0.041)			(0.029)	(0.045)
Spillover in low saturation			0.012	0.019			0.015	0.016			0.024	0.020
			(0.018)	(0.020)			(0.018)	(0.020)			(0.015)	(0.015)
Spillover in low saturation × Higher malfeasance			-0.012	-0.042			-0.017	-0.031			-0.074**	-0.080**
			(0.028)	(0.034)			(0.028)	(0.034)			(0.034)	(0.038)
Observations	13,251	13,251	13,251	13,251	13,251	13,251	13,251	13,251	13,251	13,251	13,251	13,251
R^2	0.56	0.62	0.60	0.69	0.56	0.62	0.59	0.69	0.58	0.64	0.61	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	0.18	0.18
Control outcome std. dev.	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
Test: null effect of Facebook ads with higher malfeasance (p value, 2-sided)	0.445	0.826			0.718	0.902			0.098	0.071		
Test: null effect of spillover with higher malfeasance (p value, 2-sided)	0.723	0.426			0.756	0.535			0.080	0.026		
Test: larger effect of Facebook ads in high (vs. low) with lower malfeasance (p value, 1-sided)			0.030	0.020			0.073	0.040			0.217	0.012
Test: larger effect of Facebook ads in high (vs. low) with higher malfeasance (p value, 1-sided)			0.038	0.550			0.174	0.676			0.460	0.232
Test: larger effect of spillovers in high (vs. low) with lower malfeasance (p value, 1-sided)			0.007	0.015			0.020	0.022			0.261	0.042
Test: larger effect of spillovers in high (vs. low) with higher malfeasance (p value, 1-sided)			0.996	0.693			0.032	0.649			0.478	0.211
Interactive covariates		✓		✓		✓		✓				

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable, randomization block fixed effects, and an interaction between treatment. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; amount of FISM funds received; 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/Q2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses.* denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests. The one-sided coefficient tests at the foot of the table test whether the effect of high saturation is larger in a given direction in high saturation than in low saturation municipalities.

Table A17: Effect of Facebook ads spillovers on precinct-level municipal incumbent party vote share (share of registered voters), by quartile of the sample irregularities distribution and access to 3G+ cell phone coverage

	Incumbent party vote (share of registered voters)	
	(1)	(2)
Spillover	0.034*	
	(0.018)	
Spillover × Q3	0.025	
	(0.045)	
Spillover × Q4	-0.098**	
	(0.045)	
Spillover × 3G+	-0.007	
	(0.020)	
Spillover × Q3 × 3G+	-0.066	
	(0.050)	
Spillover × Q4 × 3G+	0.067	
	(0.045)	
Spillover in high saturation		0.056
		(0.037)
Spillover in high saturation × Q3		0.074
		(0.070)
Spillover in high saturation × Q4		-0.109**
		(0.044)
Spillover in high saturation × 3G+		0.019
		(0.038)
Spillover in high saturation × Q3 × 3G+		-0.221***
		(0.070)
Spillover in high saturation × Q4 × 3G+		0.055
		(0.047)
Spillover in low saturation		0.024
		(0.024)
Spillover in low saturation × Q3		0.016
		(0.040)
Spillover in low saturation × Q4		-0.120**
		(0.057)
Spillover in low saturation × 3G+		-0.013
		(0.024)
Spillover in low saturation × Q3 × 3G+		-0.034
		(0.044)
Spillover in low saturation × Q4 × 3G+		0.099*
		(0.056)
Observations	8,683	8,683
R^2	0.65	0.71
Control outcome mean	0.18	0.18
Control outcome std. dev.	0.11	0.11
Mean 3G+	0.816	0.823

Notes: Each specification is estimated using OLS, and includes a lagged dependent variable, randomization block fixed effects, and an interaction between treatment. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; amount of FISM funds received; 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/Q2. All segments are weighted equally and by the inverse probability of treatment assignment. Standard errors clustered by municipality are in parentheses.* denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests. The one-sided coefficient tests at the foot of the table test whether the effect of high saturation is larger in a given direction in high saturation than in low saturation municipalities.

leverages only around 30% of the segments used for the main analysis, which is why we do not interact direct and indirect treatment with saturation, we again find little evidence to suggest that the mistargeting of Facebook ads explains our main findings.

A.14 Belief updating in the WhatsApp survey

One mechanism through which the spillover and saturation effects could influence voting behavior is via belief updating. The preceding findings that the Facebook ads induced greater support for the best-performing incumbents and potentially somewhat reduced support for the worst-performing incumbents is consistent with voters updating their relatively pessimistic prior beliefs. While underpowered by the low treatment compliance rate in the survey (see Table 3), a more direct test of the belief updating mechanism, in a context that is unlikely to activate coordination mechanisms, leverages the WhatsApp version of the treatment within our panel survey.

Although they are not statistically significant, the results in Table A19 are broadly in line with the precinct-level vote share outcomes. The 1 percentage point decrease in perceived irregularities in Q1/Q2 is broadly in line with the electoral returns data, but is small in magnitude. However, this reduced form effect would grow considerably after accounting for only 9% of voters recalling the WhatsApp message and video ad. These results thus suggest that belief updating may account for some of the observed effects on incumbent support, but are not definitive. The magnitude of the effects also suggests that other mechanisms could also be at play.

Table A18: Effect of Facebook ads spillovers on precinct-level vote share of nearby municipal incumbent parties (share of registered voters), by quartile of the sample irregularities distribution

	Vote for incumbent party in nearby experimental municipality (share of registered voters)			
	(1)	(2)	(3)	(4)
Facebook ads	0.008	-0.003		
	(0.040)	(0.026)		
Facebook ads × Q3	-0.014	0.013		
	(0.047)	(0.044)		
Facebook ads × Q4	0.024	0.090*		
	(0.052)	(0.054)		
Spillover	-0.011	-0.015		
	(0.044)	(0.034)		
Spillover × Q3	0.052	0.043		
	(0.062)	(0.052)		
Spillover × Q4	0.088	0.106*		
	(0.066)	(0.060)		
High saturation			0.022	0.013
			(0.041)	(0.026)
High saturation × Q3			-0.002	0.004
			(0.045)	(0.044)
High saturation × Q4			-0.006	0.097*
			(0.054)	(0.051)
Low saturation			-0.024	0.013
			(0.045)	(0.030)
Low saturation × Q3			0.039	0.014
			(0.069)	(0.061)
Low saturation × Q4			0.117*	0.132**
			(0.066)	(0.062)
Observations	642	642	642	642
R^2	0.16	0.46	0.17	0.57
Control outcome mean	0.16	0.16	0.16	0.16
Control outcome std. dev.	0.11	0.11	0.11	0.11
Test: null effect of Facebook ads in Q3 (p value, 2-sided)	0.825	0.744		
Test: null effect of Facebook ads in Q4 (p value, 2-sided)	0.379	0.036		
Test: null effect of spillover in Q3 (p value, 2-sided)	0.362	0.439		
Test: null effect of spillover in Q4 (p value, 2-sided)	0.117	0.064		
Test: larger effect in high (vs. low) saturation in Q1/Q2 (p value, 1-sided)			0.128	0.503
Test: larger effect in high (vs. low) saturation in Q3 (p value, 1-sided)			0.463	0.566
Test: larger effect in high (vs. low) saturation in Q4 (p value, 1-sided)			0.947	0.727
Interactive covariates		✓		✓

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; amount of FISM funds received; 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/Q2. 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.* denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests. The one-sided coefficient tests at the foot of the table test whether the effect of high saturation is larger in a given direction in high saturation than in low saturation municipalities.

Table A19: Effects of receiving Facebook ads via WhatsApp on perceptions of municipal incumbent party irregularities among panel survey respondents

	Perceived % irregularities (endline)		
	(1)	(2)	(3)
WhatsApp ad	1.667	2.314	2.350
	(1.204)	(1.529)	(1.678)
WhatsApp ad × Q3		-0.455	1.846
		(3.937)	(4.462)
WhatsApp ad × Q4		-2.221	-3.900
		(2.781)	(3.318)
Perceived % irregularities (baseline)	0.374***	0.374***	0.378***
	(0.025)	(0.025)	(0.025)
Observations	1,360	1,360	1,360
R^2	0.260	0.261	0.267
Outcome range	[0,100]	[0,100]	[0,100]
Control outcome mean	58.89	58.89	58.89
Control outcome std. dev.	24.53	24.53	24.53
Interactive covariates			✓

Notes: Each specification is estimated using OLS, and includes municipality fixed effects. Specifications including interactive covariates further include interactions between treatment conditions and the following municipal-level covariates: year of audit; amount of FISM funds received; 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. All observations are weighted by the inverse probability of treatment assignment. Observations from municipalities where no WhatsApp treatment were distributed are excluded. Robust standard errors are in parentheses. * denotes $p < 0.1$, ** denotes $p < 0.05$, *** denotes $p < 0.01$ from two-sided t tests.