

How Social Media Shapes Political Attitudes Through Substitution

Nicholas Kuipers*

Abstract

I study how social media shapes political attitudes in Indonesia. In a field experiment among a nearly nationally representative sample ($N = 1,502$), I deactivated five social media platforms for thirty days, finding the intervention reduced perceived corruption, possibly because treated participants had worse access to salient information about incumbent performance. Deactivation also unexpectedly increased religious chauvinism. Follow-up evidence suggests a time-use substitution mechanism: participants blocked from accessing social media reported greater involvement in religious organizations and more exposure to politicized sermons. I corroborate these findings with independently collected survey data, showing that respondents in villages with plausibly exogenous mobile internet coverage report lower satisfaction with government anti-corruption efforts, increased religious tolerance, and also reduced participation in religious organizations. Overall, the findings suggest that social media's effects on political attitudes are highly context-dependent and may chiefly reflect that which it crowds out.

*Assistant Professor, Department of Politics, Princeton University

For discussions and feedback on the design of this experiment, I thank Vin Arceneaux, Nejla Asimovic, Thad Dunning, Andy Guess, Deni Irvani, Saiful Mujani, Jonathan Mummolo, Lucas Ophoff, Grigo Pop-Eleches, Alexander Sahn, Darin Self, Jake Shapiro, and Seth Soderborg. I especially thank the staff at SMRC for their assistance in implementing the experiment, especially Saidiman Ahmad, Irwan Amrizal, and Dudi Herlianto. I also thank seminar audiences at the ESOC, Emory University, University of Michigan, Stanford University, and SEAREG. This research was reviewed and approved by the Princeton University Institutional Review Board (Protocol #18430). An anonymized version of the pre-registration can be found on the OSF directory ([here](#)). I acknowledge funding support from Princeton University. Claude Code (Opus 4.6) and ChatGPT (5.2) were used in the copyediting of this manuscript and in the construction of the code to produce the analyses contained in this paper. All errors are mine.

1 Introduction

Social media is often blamed for the sorry state of contemporary political discourse (Sunstein 2018; Tucker et al. 2017; Twenge et al. 2020). Yet recent experimental studies, in which participants deactivate their social media accounts, generally find null or modest effects on a suite of political behaviors and attitudes (Allcott et al. 2020, 2024; Arceneaux et al. 2024; Nyhan et al. 2023). In the United States, deactivating Facebook reduced news consumption and political knowledge but had little effect on polarization (Allcott et al. 2020).¹ Similar results appear in France and Bosnia and Herzegovina (Arceneaux et al. 2024; Asimovic et al. 2021). Likewise, removing Facebook’s algorithmic prioritization of salacious content had no effect on political outcomes (Guess et al. 2023). Synthesizing this literature, a recent working paper by Arceneaux, Zengin and Ladd (2025, 2) conclude that “if we do live in a polarizing, distrustful, depressing dystopia, social media provide the canvass on which our nightmare society unfolds, as opposed to its source.”

This paper takes an alternative perspective by building on research on media displacement to develop a framework that focuses on social media as a time-use substitute (Gentzkow 2006; Prior 2007). That is: while social media may or may not exert direct effects, its political consequences depend on how users would have counterfactually spent their time. The emerging experimental consensus crystallizing around minimal effects may thus reflect the character of the menu of available substitutes in the contexts in which most studies have been conducted. In these settings, chiefly the United States and Western Europe, scholars have documented a long-run rise in political polarization alongside a coeval erosion of social capital (McCarty 2019; Putnam 2000). These trends predate the advent of social media, meaning that time spent online displaced activities that were already isolating and politically polarizing, such as solitary leisure or partisan cable television (Campante, Durante and Tesei 2022; Stevic et al. 2021). Deactivating platforms in these settings thus routes users back toward equally corrosive alternatives, potentially washing out any observable effects. Across much of the Global South, however, social media arrived when offline associational life was comparatively robust, such that the alternatives to time spent online may be in-person activities with political consequences distinct in direction and magnitude from the substitutes common in Western settings.

I develop this argument with a combination of experimental and observational evidence from Indonesia—the world’s fourth most populous country and a context in which social media adoption has surged over

¹ See, also, Levy (2021) for an analysis of random variation in exposure to either in- or out-partisan news sources on Facebook.

the past decade, displacing both broadcast media and traditional face-to-face communication. The main empirical exercise is an original large-scale field experiment in Indonesia in which I recruited a nearly nationally representative sample of 1,502 adult volunteers to participate in a study in which I deactivated five social media platforms for both individual users and households for a 30-day period. For the experiment, participants were randomly assigned participants to one of three conditions: (i) individual deactivation of five platforms (Facebook, Instagram, TikTok, YouTube, and Twitter/X) for thirty days; (ii) household-level deactivation of the same platforms, in which both the respondent and other household members were incentivized to deactivate; or (iii) a status quo control group. All participants received large financial incentives, equivalent to approximately 15% of median monthly household income. To measure compliance, I developed a custom screen-time tracking application and monitored participants throughout the intervention.

I establish two central findings based on the experimental design. First, social media deactivation reduced perceptions of government corruption and secrecy: respondents cut off from social media were less likely to believe that public officials were concealing information. Consistent with a mechanism in which this shift reflects deactivation worsening participants' access to an information environment where salient political news circulates, the intervention also reduced participants' correct recall of the truthfulness of some contemporaneous news stories, although the overall index on this outcome is unchanged. Surprisingly, I detect no effect on overall satisfaction with government performance, suggesting an attenuated link between performance and satisfaction, consistent with recent work on electoral accountability (Dunning et al. 2019). Second, social media deactivation unexpectedly increased religious chauvinism along several dimensions. Respondents assigned to individually deactivate expressed greater support for violence in response to religious grievance and also became more likely to define Indonesian national identity in explicitly religious terms, rejecting the more conventional, religiously inclusive formulation. I also document an increase in generic measures of intolerance toward religious outgroups among those whose households were assigned to deactivate, although these results are more sensitive to the econometric specification.

To probe the mechanisms driving the effect on religious chauvinism, which runs counter to my pre-registered hypothesis, I fielded a follow-up survey roughly six weeks after the intervention ended among participants in the original study. Because this exercise was designed after observing the experimental results, it was explicitly aimed at evaluating my preferred post-hoc interpretation focusing on substitution. Respondents assigned to deactivate reported greater participation in offline religious organizations and a

higher likelihood of exposure to politically inflected sermons at mosques during the study period. They also reported stronger feelings that their religion was under threat. Taken together, these patterns are consistent with a substitution mechanism: social media use may crowd out time spent in offline institutions that foster exclusionary attitudes, such that blocking access to social media shifts individuals toward settings that heighten the salience of religious distinctions.

To bolster confidence in this unexpected result, I also conducted an additional empirical exercise to assess its robustness. Specifically, I assembled a large-scale harmonized dataset of survey responses ($N = 203,830$) collected across 152 nationally representative face-to-face surveys conducted from 1997–2024. For each survey, I identified questions closely analogous to those measured in the experiment. I then exploited village-level variation in mobile internet coverage driven by terrain ruggedness and proximity to major cities, drawing on the instrumental variables method developed by Olken (2009) in the same context to study the effects of television coverage. This analysis yields directionally consistent estimates with those obtained from the experiment (albeit in reverse): greater mobile internet access is associated with higher tolerance toward religious minorities, lower rates of membership in religious organizations, and lower satisfaction with government efforts to tackle corruption. Taken together, these results mirror the experimental evidence: access to social media appears to promote religious tolerance by reducing participation in chauvinism-inducing organizations while also increasing exposure to politically damaging information that reduces satisfaction with incumbents.

There are two main contributions of this paper. The first is methodological, with several features of the experimental design warranting further comment. For one, to my knowledge, this is the first social media deactivation experiment conducted on a nearly nationally representative sample (see Kuipers, Irvani and Mujani (n.d.) for the companion paper looking at mental health outcomes). It seems likely that treatment effects may vary with baseline patterns of platform use, thus making estimates from convenience samples difficult to generalize.² The study was also implemented independently of the platforms themselves, which allowed me to incentivize deactivation across five major platforms rather than only those linked to a corporate partner (Allen and Tucker 2025; Piccardi et al. 2025). A positive knock-on feature of this design is that it reduces the scope for users to substitute from one blocked platform to another, thereby enabling a clearer capture of the effect of social media at large. Finally, the experiment included a household-level treatment arm, in which other household members were also encouraged to deactivate. The design addresses an

² For a framework outlining this concern, see Torreblanca (2025).

important challenge in studying social media: interference and spillovers within offline networks. More generally, it points to the value of assigning treatment at a level higher than the individual when studying the effects of social network technologies (Bowers, Fredrickson and Panagopoulos 2013; Egami 2021).³

The second contribution is theoretical. I contribute to the growing consensus that the political effects of social media are inherently relative: platforms do not operate in a vacuum, but displace other information sources and social settings (Djourelouva, Durante and Martin 2025; Asimovic, Nagler and Tucker 2023). In this respect, the paper builds most directly on Asimovic et al. (2021), who likewise show that deactivating social media in Bosnia and Herzegovina worsens intergroup tolerance. There, the authors are agnostic on whether their results reflect a direct effect from social media in boosting tolerance or an indirect effect from substitution to intolerance-inducing offline activities. By directly documenting how respondents reallocated their time, the present study identifies substitution in time-use among those blocked from social media as a central mechanism that helps explain why deactivation increases religious chauvinism through a reallocation of time to intolerance-boosting activities. Where religious institutions organize everyday community life, as they do in Indonesia (Mujani and Liddle 2009; Pepinsky, Liddle and Mujani 2018), social media may not function as an engine of polarization. Instead, its arrival may partially substitute for the sorts of offline settings that are themselves more parochial and which propagate closed worldviews. In other words, the effects of social media depend crucially on the offline environment it displaces.

2 Two Theories of Social Media’s Political Effects

This study was designed around a set of hypotheses specified before the experiment was fielded and before I observed the results.⁴ The experiment and analysis was implemented in line with the pre-analysis plan, although in the Supplemental Appendix (SA), I outline several minor deviations. The results of the experimental intervention reveal a pattern that only partially aligns with the *ex ante* predictions from the pre-analysis plan. In what follows, in the interest of transparency, I proceed in two steps. First, I lay out the pre-registered hypotheses and the theoretical logic motivating those expectations. Then I introduce a *post-hoc* framework, which I use to interpret the unexpected results, and which guides the additional empirical exercises reported in subsequent sections. I present this sequence explicitly to make clear what was anticipated at the design stage and what emerged as an explanation in response to the evidence.

³ Indeed, see Amar et al. (2025) for an example of a successful anti-misinformation campaign targeting higher units of analysis (i.e., a classroom).

⁴ An anonymized version of the pre-registration can be found on the OSF directory ([here](#)).

2.1 Pre-Registered Hypotheses: Social Media’s Direct Effects

The pre-registered hypotheses were rooted in what I term a “direct-effects” view of social media. In this perspective, social media platforms’ effects on political attitudes operate primarily through the information they deliver to users. For critics of social media, information delivered via these platforms is generally thought to be algorithmically boosted exposure to inflammatory political content that may heighten salience of group identities, thereby driving polarization (Sunstein 2018; Bail et al. 2018). Below, I develop this intuition, focusing on three genres of attitudes as main outcomes.

H1a: Deactivation decreases knowledge of contemporaneous political news

Prior research into the effects of social media on information acquisition is ultimately mixed. In Allcott et al. (2020), the authors conduct a deactivation experiment, focusing on Facebook around the 2018 U.S. midterm elections, finding that the intervention reduced news consumption and factual political knowledge while improving subjective well-being. Inducing research participants to follow news on social media increases current-affairs knowledge and improves truth discernment (Altay, Hoes and Wojcieszak 2025). However, Allcott et al. (2024) report no statistically significant effect on an overall knowledge index in a similar experiment conducted around the 2020 U.S. general election, though secondary analyses suggest declines in general news knowledge alongside reduced exposure to misinformation while off Facebook. And, similarly, a recent experimental study in Brazil in which users’ access to WhatsApp was deactivated in advance of the presidential election found that the intervention reduced exposure to misinformation but did not affect belief in false stories (Ventura et al. Forthcoming)

These muted experimental findings exist alongside a large literature on social media and misinformation. Here, observational evidence suggests that, for many users, social media is a conduit for the acquisition of erroneous rather than factual information, especially in democratic settings such as Indonesia where educational quality may not provide an antidote to users’ susceptibility to false claims (Mujani and Kuipers 2020). In the United States, however, recent research has revealed that concerns over misinformation may be overstated (Guess, Nagler and Tucker 2019; Guess, Nyhan and Reifler 2020), although there is evidence that false news spreads faster than true news on social media, crowding out factually correct stories (Vosoughi, Roy and Aral 2018). The effects of social media deactivation on knowledge of contemporaneous news depends on the balance of these countervailing forces.

H1b: Deactivation increases satisfaction with government

Related to information acquisition is the question of how social media usage affects users' evaluation of incumbents. Looking at the staggered rollout of 3G connectivity, for example, Guriev, Melnikov and Zhuravskaya (2021) document declines in government satisfaction across a global sample of survey respondents, plausibly mediated by the expanded availability of information about corruption and poor performance. Similarly, looking at U.S. survey data, social media use appears to polarize trust judgments across partisan lines, with lower trust among out-partisans (Klein and Robison 2020). Observational work using EU data finds that heavier social media use is associated with lower trust in EU institutions, especially in regions with faster broadband (Kiratli 2023). The roll-out of mobile internet in South Africa increased turnout and led to declines in incumbent support (Donati 2023). Bowles, Marshall and Raffler (2024) report evidence from a field experiment in Uganda in which participants were given access to social media, finding a similar decline in incumbent support, although the results reverse in proximity to the election.

Taken together, and in light of the expectations outlined in the hypothesis concerned with information acquisition, these strands suggest that social media may serve to expose users to criticism and negative political information. These stories are likely to target incumbents whose performance will be more heavily scrutinized. As such, deactivation of social media platforms ought to raise satisfaction with government by reducing participants' likelihood of being exposed to such information. I therefore pre-registered the hypothesis that deactivation would increase satisfaction with government and related evaluations of political leaders.

H1c: Deactivation decreases outgroup polarization and animus

One of the most prominent critiques of social media is that it heightens the salience of outgroup distinctions and amplifies hostile political discourse, thereby fueling polarization and intergroup animus. Yet experimental evidence from deactivation studies in Western contexts offers only mixed support. The Facebook deactivation experiment conducted during the 2018 U.S. midterm elections found only modest reductions in affective polarization (Allcott et al. 2020), whereas the experiment conducted during the 2020 U.S. general election reports near-zero average effects on polarization indices (Allcott et al. 2024). And, strikingly, Asimovic et al. (2021) uncover negative (rather than positive) effects of social media deactivation on intergroup feeling thermometers in Bosnia and Herzegovina.

At the same time, a robust observational literature links online activity to intergroup hostility (although

see Samet, Arriola and Matanock (2024) for a counter example in Myanmar). In Germany, temporary outages that reduced Facebook exposure are associated with fewer anti-refugee hate crimes in areas with more far-right users (Müller and Schwarz 2021). In the U.S., counties with higher early Twitter adoption have been linked to higher numbers of hate crimes Müller and Schwarz (2023). And in Russia, city-level variation in social media adoption is associated with greater expression of xenophobic attitudes (Bursztyn et al. 2019).

My pre-registered hypothesis aligned with the observational evidence. Specifically, I anticipated that social media deactivation would improve intergroup tolerance in Indonesia, where the dominant cleavage animating Indonesian politics for the past thirty years has been between secular nationalists, on the one hand, and Islamist and Islam-inflected parties, on the other hand (Warburton 2020). The logic of this hypothesis was further motivated by Indonesia’s documented deterioration in intergroup relations during the democratic era. Recent work on the topic, by studying the events surrounding the 2017 Jakarta gubernatorial election, has found that online “grievance entrepreneurs” were believed to have played a central role in heightening the salience of religious cleavages in the violent protests that rocked Jakarta for months in calling for the ouster of a Christian governor falsely accused of blasphemy (Mietzner, Muhtadi and Halida 2018; Soderborg and Muhtadi 2023). It thus seemed plausible that deactivation would reduce animus along salient ethnic or religious lines.

2.2 A Post-hoc Framework: Substitution and the Reallocation of Social Life

As I will show, the experimental results only partially comport with the expectations outlined above. In particular, I observe that deactivation reduces perceived government corruption and secrecy—broadly consistent with my first and second hypotheses. This drop in perceptions of corruption is likely related to the finding that participants whose social media platforms were deactivated also reported a lower probability of correctly identifying some (but not all) contemporaneous news stories. However, I simultaneously observe increases several measures of religious chauvinism and intolerance, an effect that runs counter to my pre-registered expectations. These findings were robust, appearing across several related outcomes (e.g., religious intolerance, support for religious violence, and conceptions of Indonesian identity). To explain these results, I develop a post-hoc explanation centered on substitution.

The key premise of the substitution framework is that social media use occupies users’ time that would have been spent elsewhere. Indonesian social media users in the sample report spending roughly three

hours per day on the blocked platforms. When platforms are blocked, users do not merely consume less information; their time-budget constraint loosens by a large margin and thus enables a reallocation toward other activities. The removal of access to social media likely creates a time void that users fill with whatever substitutes their environment provides. The political consequences of the intervention then depend on the political and social character of whatever users choose to substitute. Call this the *opportunity cost* of time spent on social media. Below, I emphasize two such activities and highlight their relevance for political attitudes.

First, it seems very likely that this reallocation goes towards other forms of more conventional media. This logic echoes earlier work on media displacement (Prior 2007). The introduction of television crowded out radio listening, newspaper reading, and social visiting (Gentzkow 2006). The arrival of the internet broadly, and social media specifically, displaced conventional broadcast television and newspaper consumption (Djourelouva, Durante and Martin 2025). Consistent with this possibility, newspaper consumption in Indonesia has plummeted over the last twenty years: in 2003, 35% of Indonesian adults reported reading the newspaper at least once a week. In 2023, that figure had dropped to 3% (Kuipers 2026). This broadly mirrors trends observed across the Global North, as well (Hopkins 2018; Moskowitz 2021). The consequences of this substitution are likely to be large. The upfront costs associated with cable news—or the economies of scale necessary to sustain print newspaper distribution networks—mean that such news services are prone to be captured by the wealthy whose oligarchic interests are likely to be communicated via these traditional outlets (see Tapsell (2017) for a description of these dynamics in Indonesia).

The second activity which social media may have displaced concerns offline associational life. Early work showed that the arrival of internet access reduced face-to-face socializing in its early years in the United States (Kraut et al. 1998). Work by Olken (2009) in the Indonesian context looked at the roll out of broadcast television and radio across Indonesian villages in the early 2000s, finding modest drops in social capital and declines in associational life. It seems plausible this finding also extends to the arrival of mobile internet and social media. Social media is highly addictive (Bursztyn et al. 2025), after all, and the nature of its consumption depends on its users being atomized from one another—features that would appear likely to displace offline interaction.

The character of these opportunity costs is likely to vary across contexts. Consider, first, the United States and Western Europe, where most deactivation experiments have been conducted. There, the menu of offline substitutes is often dominated by other polarizing media (e.g., cable news) or relatively individ-

ualized leisure activities (e.g., bowling alone). Logging off social media may therefore reproduce a similar informational environment, helping explain why existing experiments in these settings have found only modest political effects. But the pattern may differ in societies in the Global South, where offline associational life is more robust, and where activities may be organized through institutions with prejudiced ideological content.

Indonesia is a case where these substitution effects may be especially consequential. Cross-national evidence from the World Values Survey indicates that, until recently, Indonesia had among the highest rates of weekly religious attendance in the world—roughly three times the OECD average and well above other ASEAN countries. Offline religious participation, including Islamic study groups (*pengajian*) and mosque-based community activities, structures everyday social life in local Indonesian communities (Mujani and Liddle 2009). But these institutions may also strengthen in-group solidarity, heighten perceptions of religious threat, and transmit narratives that privilege religious boundaries over inclusive civic conceptions of national belonging. If deactivation shifts time toward these settings, it could increase religious chauvinism and intolerance.

Figure 1—Relationship between offline associational life and religious intolerance prior to social media

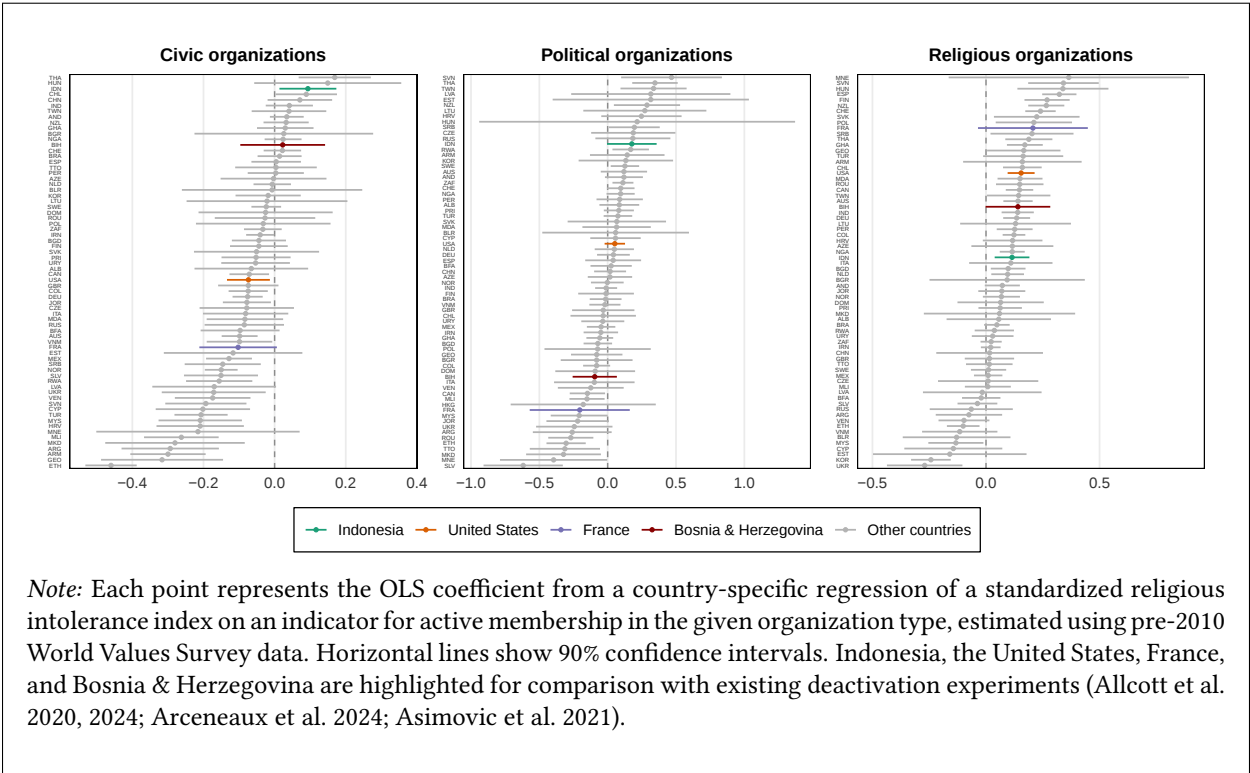


Figure 1 provides suggestive cross-national evidence consistent with this possibility. One of the key parameters of the substitution framework is the relationship between offline associational life and intolerance. In places where the relationship is positive, social media may harbor a negative effect on intolerance. But in places where the relationship is negative, the role of social media may reverse. To establish benchmarks for these expectations, I draw on World Values Survey data, subsetting to surveys conducted prior to 2010, before the widespread diffusion of social media across the developing world, using the data to estimate the country-level association between active participation in civic, political, and religious organizations and a standardized (z -score) index of religious intolerance. Across most countries, unsurprisingly, there exists a positive relationship between active membership in religious organizations and religious intolerance, including for those contexts in which social media deactivation experiments have been conducted up until this point, which are highlighted on the figure.

Among the roughly 70 countries with available data, however, Indonesia is an outlier in terms of the relationship between active participation in civic and political organizations and religious intolerance. Here, respondents in Indonesia exhibit a large, positive association between civic organizational participation and intolerance. In Western democracies, civic organizations—sports clubs, environmental groups, humanitarian associations—are conventionally understood as sites of “bridging” social capital that foster cross-cutting ties across ascriptive lines and thus reduce prejudice (Putnam 2000). In Indonesia, by contrast, even nominally civic organizations appear embedded in social networks that induce greater religious intolerance. The United States and France, where prior deactivation experiments have found modest effects on polarization, sit near the center of the distribution, consistent with an environment where offline substitutes carry less distinctive ideological content.

3 Research Design

I employ three complementary empirical strategies to evaluate these claims: (1) a three-arm deactivation experiment, (2) a retrospective follow-up survey among the original study participants to probe mechanisms, and (3) an instrumental variables analysis using geographic variation in mobile internet coverage to corroborate the results in a distinct sample.

3.1 Study #1: Deactivation Experiment

A total of 1,502 index respondents were randomly assigned to one of three conditions targeting a 400:400:700 allocation via sample random assignment at the household level.⁵ I define N_0 as the number of respondents in each condition at baseline, N_1 as the number of respondents at endline, and N_2 as the number who completed the follow-up survey. The control group ($N_0 = 402$, $N_1 = 383$) continued using social media as usual but received the same incentive as treated respondents to avoid confounding. The individual deactivation group ($N_0 = 399$, $N_1 = 367$) was incentivized to deactivate five platforms for thirty days using the StayFree application, which blocks access to selected apps and also eliminates access through web browsers. Enumerators ensured that the application was downloaded onto all of the respondent’s devices (e.g., tablet, cell phone, and laptop). Compliance was monitored via an additional tracking application which was available for Android devices (which comprised 95% of our sample). The compliance monitoring app was not available for iOS, and so for those participants on Apple devices, compliance was monitored via daily screen-time screenshots delivered to a team of research assistants. The household deactivation group ($N_0 = 701$, $N_1 = 661$) replicated the individual protocol but extended eligibility and incentives to all consenting adult household members, with the aim of reducing within-household informational spillovers that could attenuate individual-level treatment effects. In the event, 32.5% of households assigned to this condition had additional participants deactivate social media. All analyses focus on the index respondent. Figure A1 shows the CONSORT diagram for the experiment.

Incentives were calibrated via a 40-person pilot in Greater Jakarta in October 2025, discovering study entry to be inelastic beyond 500,000 IDR (approximately US\$30, or 14% of median monthly household income), which I adopted as the benchmark in the full study. I further randomized incentives at 400,000, 500,000, or 600,000 IDR to study compliance as a function of incentive level to estimate the elasticity of respondents’ social media addiction as part of a separate project. The structure of the payments was such that participants received 20% upon completing the baseline survey and the remaining 80% upon completing the endline, delivered via e-wallet (OVO, GrabPay, or GoPay). Starting in November 2025, participants were recruited face-to-face by trained enumerators. Recruitment continued through December 17, 2026. The deactivation period was 30 days for all respondents, meaning that the experimental intervention started and ended on different days for different respondents, depending on when they were recruited. The endline

⁵ The household arm was intentionally oversampled to accommodate anticipated noncompliance given the higher burden of coordinating deactivation across multiple adults.

survey was administered online. The overall recontact rate was 93.9% with no differential attrition across experimental conditions. I group the attitudinal outcomes into conceptually related families: corruption and transparency, government evaluations, religious intolerance, political violence, national identity, and political knowledge. The descriptive statistics and the question texts can be found in SA Tables A1 and A4, respectively. Then, I summarize each “family” of outcomes using a control-group standardized index (Kling, Liebman and Katz 2007), with baseline control-group values applied to both waves.

For estimation, I focus on intent-to-treat (ITT) effects via OLS:

$$Y_{i,1} = \gamma_0 + \gamma_1 T_i + X_i' \theta + \delta_d + \epsilon_i \quad (1)$$

where I estimate ITT effects separately for T1 and T2, benchmarked against the control condition. Specifically, for the former, T_i is a treatment indicator that takes a “1” if respondent i is assigned to T1 and a “0” if respondent i is assigned to control, and missing otherwise. For the latter, T_i is a treatment indicator that takes a “1” if respondent i is assigned to T2 and a “0” if respondent i is assigned to control, and missing otherwise. Otherwise, $Y_{i,1}$ is the endline index, δ_d is a district fixed effect term, and X_i is a vector of pre-registered controls including age, gender, religion, and education. I calculate heteroskedasticity-robust standard errors. Because compliance was imperfect, I complement ITT estimates with local average treatment effects (LATE) via 2SLS, instrumenting actual deactivation with randomized assignment.⁶ I present the LATE estimates in SA Section C.1.1. Consistent with the pre-analysis plan, I present both conventional p -values and false-discovery rate (FDR) adjusted q -values, in which I account for multiple hypothesis testing within genres of outcomes (Benjamini and Hochberg 1995). As discussed above, key design elements and outcome definitions were pre-registered. Throughout the manuscript, I make clear which outcomes and estimates are not pre-registered and should thus be treated as exploratory.

3.2 Study #2: Follow-Up Mechanism Survey

To probe mechanisms, approximately six weeks after the conclusion of the experimental intervention (and once I had observed the main effects), I fielded a follow-up survey. Here, I recontacted participants who completed the endline survey from all three experimental arms and administered a short follow-up survey designed to probe the substitution mechanism. To boost response rates, respondents were offered a small

⁶ I define compliance as averaging fewer than 5 minutes of daily social media use on each platform over the intervention period, based on device monitoring data. This is the pre-registered compliance measure used throughout the LATE estimation. See the SA for the results under LATE estimation.

incentive (20,000 IDR, i.e., US\$ 1.50) via the same e-wallet delivery mechanism that was used before. In the end, 974 respondents completed the follow-up survey for a response rate of 69.0%. There was no meaningful differential attrition across treatment arms, as reflected in the response rates ($R_c = 70.8\%$, $R_{T1} = 66.2\%$, $R_{T2} = 69.6\%$). I report the descriptive statistics for the follow-up mechanism survey sample in SA Table A7.

The survey probed respondents about my preferred interpretation concerning time-use substitution as the mechanism that explains the results observed in the original study. So, specifically, I asked the study participants to report on time reallocation during the study period, with modules covering: participation in offline religious organizations (e.g., religious study groups (*pengajian*) and mosque-based activities), exposure to politically inflected sermons, and time spent watching television including government-aligned programming. These items allow me to test directly whether deactivated respondents substituted toward the offline institutions that the theoretical framework identifies as consequential for the observed pattern of effects. I construct control-group standardized indices, again, and estimate the effects of the intervention on these outcomes using Equation 1. It is worth emphasizing that these analyses were not pre-registered as they were part of the post-hoc substitution framework.

3.3 Study #3: Instrumental Variables Analysis

To assess whether the experimental patterns generalize beyond the study sample, I leverage geographic variation in mobile internet coverage across Indonesia using an instrumental variables approach. Specifically, I follow the strategy introduced by Olken (2009) who examines the effects of the roll-out of broadcast television by leveraging Indonesia’s mountainous terrain. I adapt this method by combining distinct data sources, showing the instruments proposed by Olken (2009) to be even stronger predictors of mobile internet coverage: (1) the standard deviation of elevation (terrain ruggedness), which proxies for the difficulty of installing cell towers in mountainous terrain; and (2) the log distance from the village centroid to the nearest major city, which captures the cost of extending network infrastructure to remote locations. Together, these variables jointly affect the cost and feasibility of network deployment and thus generate plausibly exogenous variation in coverage conditional on province and year fixed effects.

For measurement, I harmonize over 200,000 individual-level survey responses spanning 1999–2024, drawn from 152 nationally representative face-to-face surveys.⁷ Importantly, these surveys probed whether

⁷ These surveys were conducted by Saiful Mujani Research and Consulting (SMRC). SMRC is the firm that assisted in implementing the field experiment reported in this paper.

respondents used the internet or not. The final analysis includes up to 108,419 respondents for whom I was able to geocode their location and to whom the pertinent questions were asked. For the outcomes, I identify survey questions closely analogous to the experimental outcomes, focusing on religious intolerance (objection to outgroup religious activities in one’s community) and political satisfaction (evaluations of corruption, presidential approval, and democratic performance). These survey responses provide information on whether respondents use mobile internet and, crucially, their village-level location, which I use to geolocate each response. All relevant variables are summarized in SA Table A8.

I then estimate the effect of mobile internet coverage on the relevant political attitudes via a two-stage least squares system of equations:

$$\text{First stage: } \text{Internet}_{idt} = \alpha + \pi_1 \text{ElevationSD}_d + \pi_2 \text{LogDistCity}_d + \mu_p + \lambda_t + \varepsilon_b \quad (2)$$

$$\text{Second stage: } Y_{idt} = \beta_0 + \beta_1 \widehat{\text{Internet}}_{idt} + \mu_p + \lambda_t + \varepsilon_b \quad (3)$$

where i indexes individuals, d villages, b districts, p provinces, and t years; province and year fixed effects absorb time-invariant geographic and temporal confounders, and standard errors are clustered at the district level. The instruments are constructed at the village level and matched to respondents via geocoded survey locations.

The key assumption underpinning identification under an instrumental variables empirical strategy is the so-called “exclusion restriction” which holds that the only channel through which the instrument affects the outcome is through the endogenous regressor. In the present case, the exclusion restriction would be violated if other factors correlated with terrain ruggedness, distance to major city, and the attitudinal outcomes in question. One of the most likely candidates for such a violation is the level of economic development, which may plausibly and jointly affect the likelihood of internet adoption as well as, say, incumbent satisfaction. To address this concern, I adopt the strategy proposed by Olken (2009) of including a vector of controls in both stages of the estimation that condition on levels of economic development, which I proxy using share of households with electricity, driving distance to nearest major city, and village-level distance to Jakarta (See Tables A12, A13, and A14).

4 Analysis #1: Deactivation Field Experiment

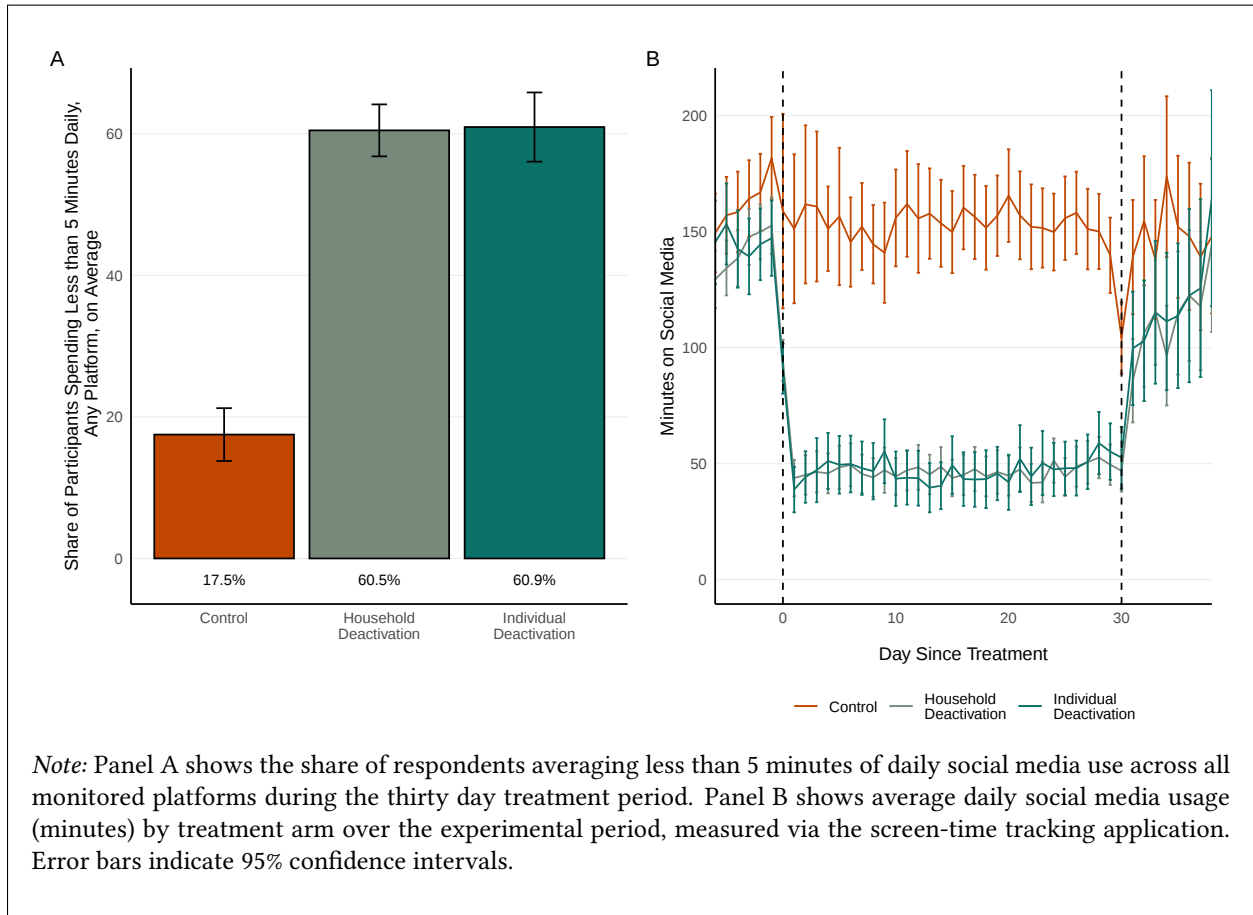
4.1 Compliance

I start by assessing participants' compliance with the deactivation experiment. Recall that I measured compliance using a bespoke application developed for Android devices (95% of the sample), while a team of research assistants manually verified compliance on a daily basis for iOS devices by screenshots of the time-use portal. Figure 2 shows strong evidence of compliance. The benchmark measure of compliance (pre-registered to use for LATE estimation) was whether a respondent spent less than 5 minutes per day on average across all five platforms. Panel A reports that approximately 61% of treated respondents in both the individual and household arms maintained average daily social media usage below five minutes across all monitored platforms during the thirty day treatment period, compared to roughly 18% of control respondents—a gap of over 40 percentage points. I detect no meaningful variation in compliance across the two treatment conditions.

Panel B tracks daily social media usage over the course of the study. Prior to the intervention, all three groups averaged approximately 150 minutes of daily social media use, confirming that randomization successfully balanced baseline usage. At the onset of treatment, usage in both deactivation arms dropped sharply to approximately 40–50 minutes per day and remained at that level throughout the intervention window. Control-group usage, by contrast, remained stable at baseline levels, although there appears to have been a slight decline in usage after enrollment in the study, likely a function of the “nudge” of having installed the compliance tracking app. After the treatment period concluded, usage in the deactivation arms rebounded but did not immediately recover to pre-treatment levels, suggesting some persistence in altered habits.⁸ Again, the compliance patterns are nearly identical across the individual and household treatment arms, indicating that the logistical burden of coordinating household-level deactivation did not meaningfully reduce adherence.

⁸ The apparent drop in control-group usage on the final day of tracking reflects displacement of time owing to survey completion effort on the day the endline survey was administered, rather than a behavioral change.

Figure 2—Compliance with social media deactivation



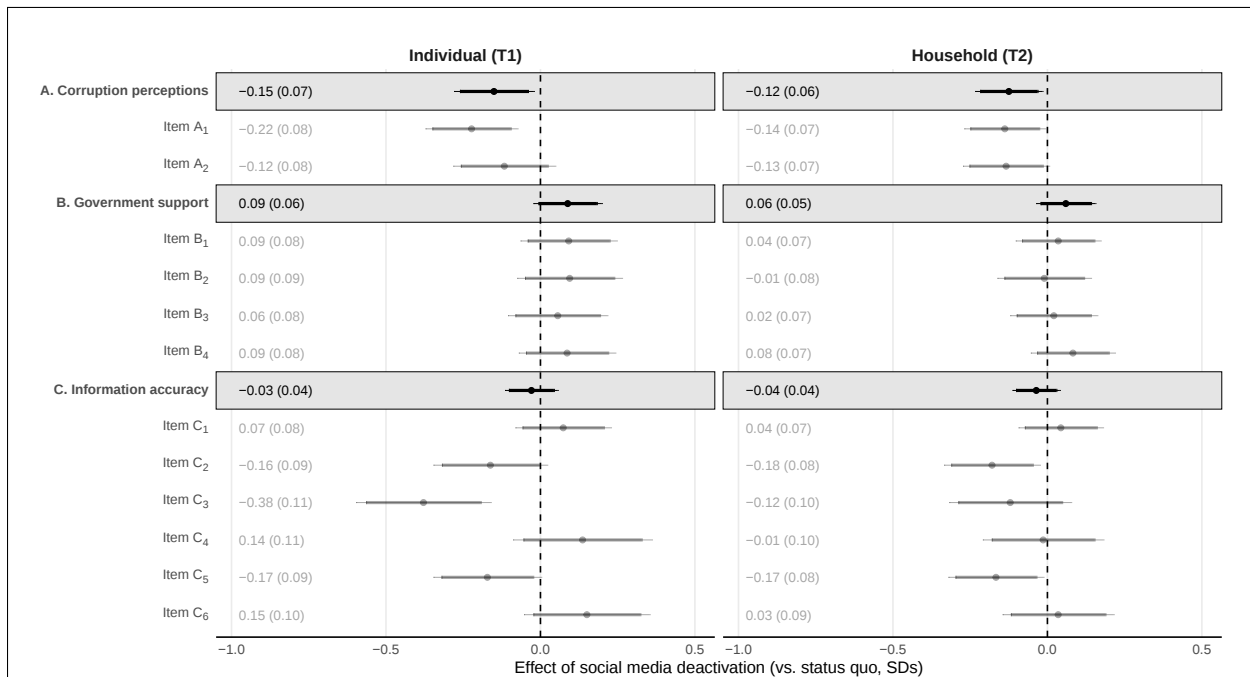
4.2 Effects on Corruption Perceptions and Political Information

I begin the analysis of the deactivation field experiment by examining outcomes related to governance. Recall that the over-arching pre-registered hypothesis was that social media deactivation would worsen participants' access to information, in turn driving up support for incumbent politicians otherwise weighed down by reports of their malfeasance. Figure 3 displays ITT estimates for the individual-deactivation arm (T1, left panel) and the household-deactivation arm (T2, right panel), reporting both summary indices (shaded rows) and their constituent items. I present 90% (thick bars) and 95% (thin bars) confidence intervals throughout.

The most consistent experimental finding concerns corruption perceptions, although it is worth emphasizing that this particular index was not pre-registered. The corruption perceptions index declines by 0.15 standard deviations in the individual arm ($p < 0.05$, $q < 0.05$) and by 0.13 SD in the household arm

($p < 0.05$, $q < 0.05$). The constituent item capturing beliefs about (A_1) whether government officials hide information from the public drives this result, declining by 0.22 SD under individual deactivation ($p < 0.01$, $q < 0.01$) and 0.14 SD under household deactivation ($p < 0.05$, $q < 0.05$). The near-identical magnitudes of the index-level estimates across treatment arms (0.15 SD vs. 0.13 SD) suggest that within-household spillovers do not appreciably attenuate individual-level treatment effects for this outcome. Respondents who were cut off from social media became less likely to believe that officials were engaged in the concealment of malfeasance, consistent with the general expectation outlined in the pre-registered hypotheses. My preferred interpretation is that social media exposes users to critical political information, including reports of official malfeasance, that were made less salient when users were deactivated.

Figure 3—Effects of social media deactivation on governance-related outcomes



Note: Point estimates are ITT coefficients with 90% (thick) and 95% (thin) confidence intervals. Shaded rows are Kling-style summary indices; unshaded rows are constituent items. Item A₁: “How often do government officials hide important information from the public?” (1–4). Item A₂: “How are corruption and bribery practices among government officials?” (1–4). Item B₁: Satisfaction with President Prabowo’s governance performance (1–4). Item B₂: Trust in the President (1–4). Item B₃: Trust in the national parliament/DPR (1–4). Item B₄: Trust in local government (1–4). Item C₁: “Indonesia’s current finance minister is Sri Mulyani” (True/False). Item C₂: “The National Police Chief has been replaced” (True/False). Item C₃: “Jakarta plans to relocate the IKJ campus to Lebak Bulus” (True/False). Item C₄: “Jakarta operates RDF Plant Rorotan to address the waste crisis” (True/False). Item C₅: “About 40% of the national education budget is allocated to the free school meals program” (True/False). Item C₆: “The North Maluku Governor is the wife of the late former Morotai regent” (True/False). Tabular regression estimates are reported in Table A9.

By contrast, the government support index shows no detectable movement in either arm, with point estimates close to zero. Approval for incumbent President Prabowo (B_1), as well as generalized trust in the presidency (B_2), national legislature (B_3), and local government (B_4) all remain statistically indistinguishable from control levels. This null result is noteworthy in the context of the findings concerning corruption: while deactivation shifted perceptions of government transparency, it did not translate into changes in evaluative judgments of government performance. The former may simply reflect participants' exposure to information, which induces first-order updating, while the latter may reflect more durable partisan or affective attachments that thirty days of reduced social media access is insufficient to move.

The final item under the umbrella of governance outcomes is the information accuracy index, which shows no statistically significant changes for participants assigned to either the individual or the household deactivation. The disaggregated news items are particularly instructive for this index, however, as respondents were quizzed about individual news stories that capture distinct streams of information. In this context, examining treatment effects at the level of individual survey items is therefore warranted. These reveal a consistent pattern of negative effects. Several items show important and instructive drops. First, participants whose social media was deactivated report a large decline in correctly identifying (C_3) a claim about government plans to relocate the IKJ campus, by a margin of 0.38 SDs under individual deactivation ($p < 0.01$, $q < 0.01$), although this effect does not replicate in the household arm. Second, I detect declines in correctly identifying as false (C_2) a claim about a recent change in the national police chief by a margin of 0.16 SDs under individual deactivation ($p < 0.10$, $q < 0.10$) and 0.18 SDs under household deactivation ($p < 0.05$, $q < 0.05$). Third, I also detect a decline in whether they accurately assessed a claim about (C_5) the central government's intention to spend 40% of its national education budget on a widely-ridiculed plan to provide free school lunches, by a margin of 0.17 SDs under individual deactivation ($p < 0.10$, $q < 0.10$) and 0.17 SDs under household deactivation ($p < 0.05$, $q < 0.05$). These item-level results are consistent with the pre-registered hypothesis (H1a) that deactivation would reduce knowledge of contemporaneous political news, even though the aggregate index lacks the precision to reject the null hypothesis. These findings are also broadly consistent with the result on the corruption perceptions index given that the stories with the largest drops in accurate recall involved government accountability and resource allocation.

4.3 Effects on Religious Chauvinism

The next set of outcomes I investigate concern religious attitudes. To be clear, in the pre-analysis plan, I pre-registered an expectation that social media deactivation would improve religious tolerance (and related outcomes). Specifically, Figure 4 displays ITT estimates across three outcome families: religious intolerance, support for religious violence, and religious nationalism. As before, not all of the indices were pre-registered. The index capturing religious intolerance was pre-registered, although the index capturing religious violence and the survey item measuring religious nationalism were not. Because the survey items capturing intolerance are religiously-specific, I subset the analyses to the attitudes of Muslims, who represent 90% of the sample.

In the individual-deactivation arm, contrary to the overarching expectation of my pre-registered hypothesis, I detect an *increase* in the index measuring support for religious violence, by a margin of 0.15 SD ($p < 0.05$, $q < 0.10$), indicating that respondents assigned to deactivate their social media accounts became significantly more supportive of the use of violence in response to perceived religious grievances. It is worth stressing that these measures likely capture an expressive rather than literal sentiment. The National Violence Monitoring System, which tracks episodes of communal violence in Indonesia, suggests that religious violence is generally rare (Kuipers, Nellis and Weaver 2021; Barron, Jaffrey and Varshney 2016).⁹ Looking at this index specifically, all three constituent items move in the same direction: justification of violence in response to (B₃) blasphemy rises by 0.21 SD ($p < 0.05$, $q < 0.05$), while the items measuring justification of violence to (B₂) fight “negative influences that damage society’s morals” (0.12 SD, $p > 0.10$, $q > 0.10$) and to (B₁) prevent electoral fraud (0.13 SD, $p > 0.10$, $q > 0.10$) are positive but not statistically significant.

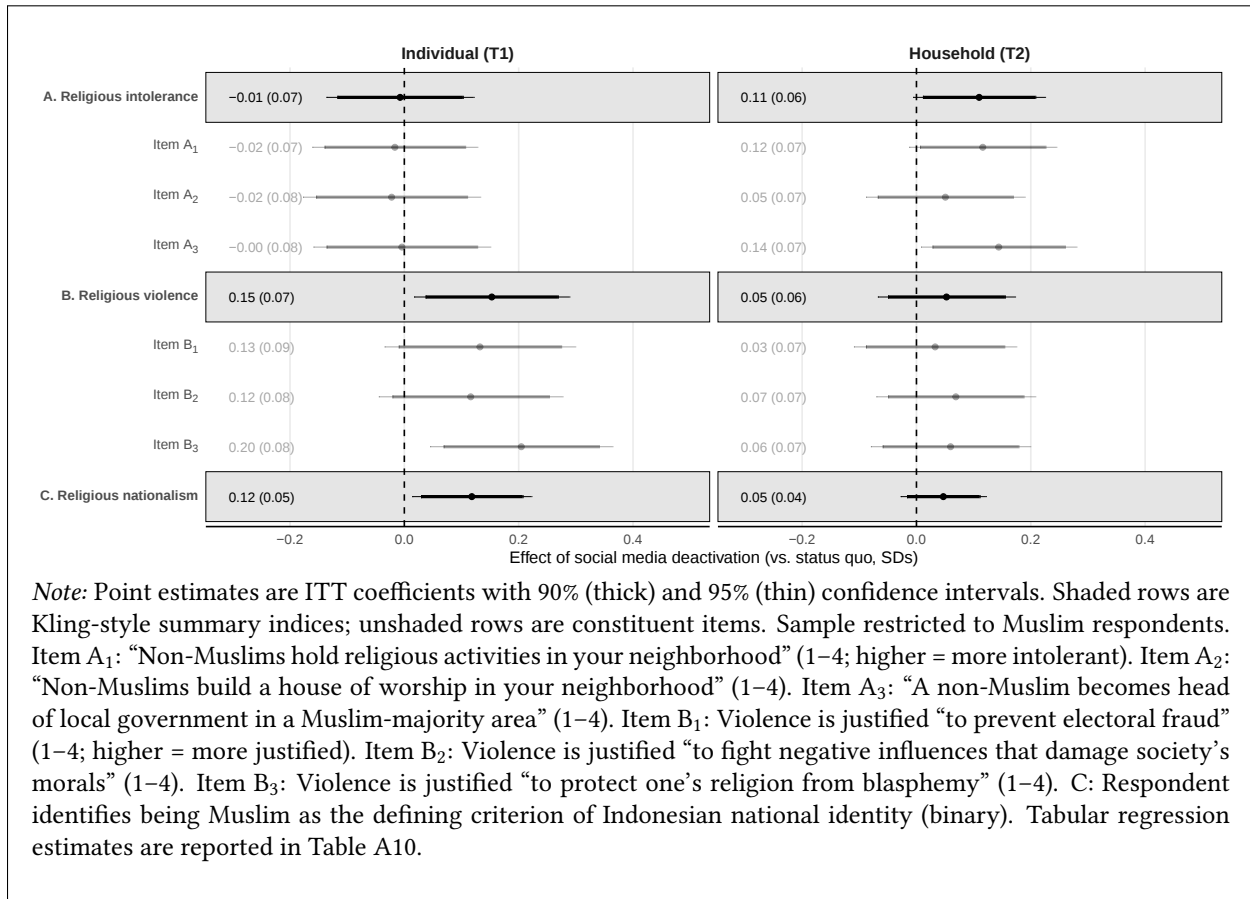
The final row of Figure 4 looks at a religious nationalism item, in which I asked respondents to identify the traits most important for being considered “Indonesian.” Recall that Indonesia is a religiously plural nation, in which approximately 12% of the population adhere to religions other than Islam. The national identity is united under the concept of *pancasila*, which is an inclusive doctrine that stipulates that all Indonesians must adhere to a monotheistic religion (without clarifying which one). I construct an indicator

⁹ A nationally representative December 2017 survey conducted by SMRC found that 0% of respondents reported having participated in an attack on a religious establishment in the past three years, and less than 1% of respondents indicated an interest in joining such an attack if the opportunity arose. The same survey asked whether respondents had participated in the last three years in a *sweeping*, a more common activity in which Islamic organizations (e.g., FPI or HTI) raid a bar or nightclub. Only 1.6% respondents in the sample reported having done so. By contrast, for a review of frequency of low-grade vigilante violence in Indonesia see Jaffrey (2023).

variable that takes a “1” if the respondent identified “Islam” as the most important factor in determining what constitutes an Indonesian, as opposed to more inclusive civic conceptions of belonging. I find this variable to rise by 0.12 SD ($p < 0.05$, $q < 0.10$) under the individual social media deactivation arm.

In the household-deactivation arm, the pattern shifts. Here, I detect no effect of household deactivation on support for religious violence, nor do I see an increase in religious nationalism, although all of the point estimates are consistent with the findings in the individual deactivation arm. Instead, however, I find that household social media deactivation increased the items capturing religious intolerance, which I measured by quizzing respondents about their willingness to accept religious outgroups (A_1) holding religious events in their neighborhood, (A_2) building houses of worship in their neighborhood, or (A_3) holding political office in their community. Specifically, I detect an overall increase of 0.11 SD ($p < 0.10$, $q > 0.10$) in the main index, although this finding does not retain statistical significance after correcting for multiple hypothesis testing.

Figure 4—Effects of social media deactivation on religious attitudes



Three features of these results merit emphasis. First, the direction of the effects on religious attitudes runs directly counter to my pre-registered hypothesis (H1c), which predicted that deactivation would *reduce* intergroup animus. Instead, removing social media appears to have increased religious chauvinism along multiple margins. Second, the effects are substantively large, and they appear across several related but distinct constructs including religious violence justification, exclusionary national identity, and (in the household arm) intolerance toward religious minorities. Third, the pattern of which arm produces which effects is itself suggestive. Individual deactivation, which leaves household members' social media use intact, produces stronger effects on religious violence and nationalism, which are more extreme outcomes that may reflect individual-level exposure to different information environments and social settings. Household deactivation does not produce effects support for violence or exclusivist nationalism, which are more extreme statements of chauvinism, but does lead to swings on generalized intolerance, which is an outcome that may be more sensitive to the interpersonal social dynamics that shift when an entire household's media diet changes simultaneously.

5 Analysis #2: A Retrospective Follow-up Survey

My preferred explanation for the results obtained in the deactivation field experiment concerns time-use substitution. In preventing participants from exhausting their free time on social media, the intervention may have induced them to allocate their time elsewhere into activities that induced the attitudinal changes I observe. To probe this possibility, I conducted a follow-up survey. Approximately six weeks after the intervention concluded, I recontacted participants from all three experimental arms and administered a short survey probing time reallocation during the study period. All questions were framed in retrospective terms, asking respondents to report on what they did during the study period. Figure 5 reports the results across three outcome families: television substitution, participation in religious organizations, and exposure to politicized Islamic content (the latter two restricted to Muslim respondents, who constitute the large majority of the sample and for whom mosque-based organizational life is most relevant).

The television substitution index is positive in both arms but statistically insignificant. Specifically, the results indicate that respondents assigned to the individual deactivation report a 0.11 SD increase in the television consumption index ($p > 0.10$, $q > 0.10$) and a 0.06 SD in the household arm ($p > 0.10$, $q > 0.10$). Among the constituent items, items A_2 and A_3 capture whether respondents reported spending more time watching television programming that portrayed the national and local governments positively,

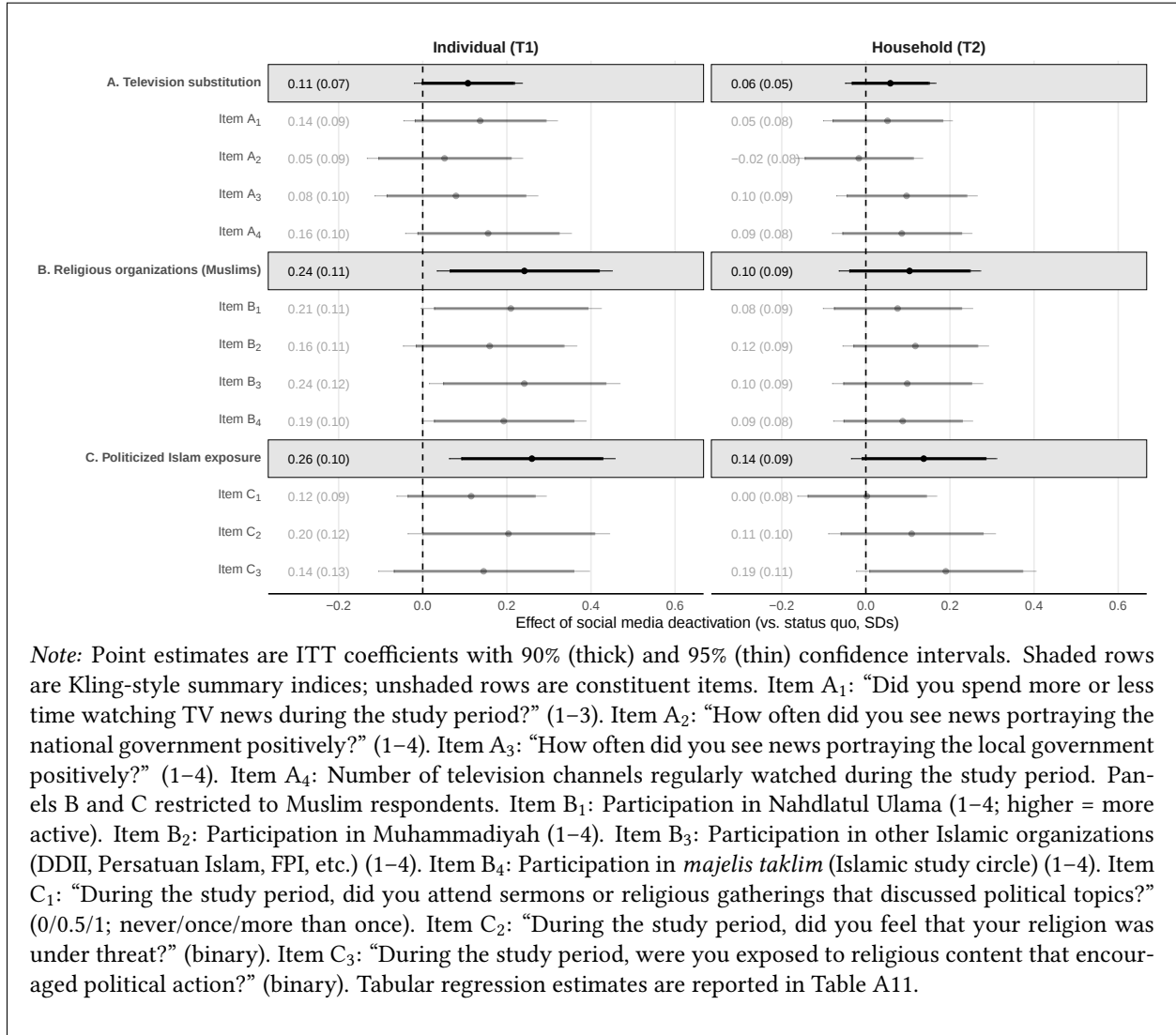
respectively, during the study period. In the household arm, the latter item shows the largest positive shift, though with wide confidence intervals reflecting considerable heterogeneity. The directional pattern is consistent with substitution toward television, but I am unable to rule out the possibility that these effects arise from chance alone.

The evidence for substitution toward religious organizations is more compelling, particularly in the individual-deactivation arm. Among Muslim respondents assigned to individual deactivation, the religious organization index increases by 0.24 SD ($p < 0.05$, $q < 0.05$). I detect positive swings in participation across multiple religious organizations, including in religious study groups (B_4 , *pengajian*) and the more hierarchical and traditionalist Nahdlatul Ulama (B_1), which is also the world's largest Muslim organization. But it is instructive that the strongest effects come from the item asking respondents about participation in other religious organizations, including DDII and FPI (B_3), the latter organization being a recently-outlawed extremist organization associated with stoking and sparking the 2017 riots in Jakarta in which an estimated 1 million people descended on the city to participate in a protest against Ahok, a Christian governor falsely accused of blaspheming the Quran (Aspinall and Mietzner 2019).¹⁰ In other words, the results point to a heightened engagement with organized Islamic life during the intervention period, including with extremist organizations.

Panel C of Figure 5 reports effects on a politicized Islam exposure index, which captures whether respondents attended sermons with political content (C_1), felt that their religion was under threat (C_2), or were exposed to religious content encouraging political action (C_3) during the study period. The index increases by 0.26 SD ($p < 0.01$, $q < 0.01$) under individual deactivation. This result is especially noteworthy: if deactivated respondents not only participated more in religious organizations but also came away from those settings with an elevated sense of religious grievance and exposure to politically charged religious messaging, this would provide a direct psychological channel linking organizational substitution to the increases in religious violence justification and exclusionary nationalism observed in the main experimental results.

¹⁰ For a fuller review of this episode, see *inter alia*, Setijadi (2017) and Mietzner (2026, ch. 6).

Figure 5—Effects of social media deactivation on time-use substitution patterns (follow-up survey)



In the household-deactivation arm, the religious organization index is not statistically significant (0.10 SD, $p > 0.10$, $q > 0.10$), and the politicized Islam exposure index is positive but statistically insignificant. This asymmetry maps onto the pattern of main effects: the household arm produced weaker effects on religious violence and nationalism than the individual arm. One interpretation is that household-level deactivation, by simultaneously removing social media from multiple family members, restructured household leisure patterns in ways that did not disproportionately channel individuals toward mosque-based activities—perhaps because household members substituted toward shared domestic activities or entertainment rather than individual religious participation.

6 Analysis #3: Instrumental Variables

The deactivation field experiment revealed that the intervention, in addition to clouding participants' perceptions of corruption, may also harbor unexpectedly negative effects vis-a-vis religious chauvinism. The results from the retrospective follow-up survey among participants confirm these results, and extend the logic by providing evidence for a substitution-style explanation. But readers may still be skeptical about the findings with respect to religious chauvinism, particularly since they cut against the pre-registered hypotheses, and because they are at odds with the conventional wisdom about social media's effects.

I therefore test whether the experimental patterns hold beyond the experimental sample and protocol. To do so, I complement the field experiment with an instrumental variables analysis in which I exploit geographic variation in mobile internet coverage across Indonesia over the past three decades. Whereas the experiment identifies the effect of *removing* social media from individual users, this analysis identifies the effect of *expanding* mobile internet access across communities—and its effect on inducing uptake of internet usage among survey respondents. It is worth stressing at the outset that this is a more conceptually expansive but nonetheless related margin to the one studied in the field experiment. Especially given the counterintuitive results on religious tolerance in the field experiment, the goal with this additional research design is to investigate whether analogous attitudinal patterns emerge at the population level and in an entirely distinct empirical setup.

The first-stage results confirm that both instruments (terrain ruggedness and log distance to the nearest major city) are strong negative predictors of social media usage on the part of individuals, consistent with the cost of network deployment being higher in rugged and remote terrain. These instruments follow the logic of Olken (2009): terrain features and proximity to urban centers that were determined long before the advent of mobile internet affect coverage through the economics of infrastructure deployment. There are nonetheless reasonable concerns over the exclusion restriction which I attempt to address by including provincial and year fixed effects. In the appendix, I also conduct an analysis in which the instrumental variables specification controls for local features thought to be jointly related to the sources of exogenous variation and the endogenous regressor, such proxies for the level of economic development.

6.1 Effects on Governance Attitudes

As before, I begin by presenting the instrumental variable estimates on the effect of individual-level internet usage on attitudes related to the quality of governance. I present the results in Table 2. Recall that the outcomes are drawn from harmonized survey data and geolocated at the village level. I first show that instrumented internet access significantly reduces respondents' evaluation of government efforts to address corruption (-0.28 , $p < 0.05$). To be clear, this is a different measure than the one captured in the field experiment, in which I directly assessed perceptions of corruption, rather than government efforts to address the issue. But both measures tap the same underlying psychological construct. I also examine presidential satisfaction (-0.45 , $p < 0.01$) and democratic satisfaction (-0.41 , $p < 0.01$), finding large declines in both, consistent with the recent work of Guriev, Melnikov and Zhuravskaya (2021) in which the author show that access to mobile internet hurts incumbent evaluations. All three point estimates are negative and precisely estimated, indicating that communities with greater predicted internet coverage express systematically lower satisfaction with political leaders and institutions.

Table 1: The Effect of Internet on Incumbent Evaluations

	First stage	Second stage outcomes		
	Internet	Pres. Satis.	Corrup. Eval.	Dem. Satis.
	(1)	(2)	(3)	(4)
Elevation SD (km)	-0.325*** (0.074)			
Dist. nearest city (log)	-0.118*** (0.013)			
Internet		-0.453*** (0.086)	-0.278** (0.129)	-0.409*** (0.072)
Observations	78,242	67,517	21,418	42,114
R-squared	0.126	-0.018	0.066	-0.030
Year FE	✓	✓	✓	✓
Province FE	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓
First-stage F	56.85	497.89	140.41	352.56

Notes: Two-stage least squares estimates of the effect of internet access on evaluations of government performance. Column (1) reports the first stage (same specification as the preceding table). Columns (2)–(4) report second-stage estimates drawn from harmonized Indonesian surveys (1999–2024): (2) satisfaction with the incumbent president (1–4; higher = more satisfied); (3) evaluation of government efforts to address corruption among government officials (1–4; higher = worse evaluation); (4) satisfaction with democracy (1–4; higher = more satisfied). All specifications include year and province fixed effects. Standard errors clustered at the district level in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

These results mirror the experimental evidence on corruption perceptions. In the experiment, de-

activating social media reduced the belief that officials hide information, implying that social media increases perceived corruption. In the instrumental variables analysis, internet access decreases satisfaction with government efforts to handle corruption, and also decreases political satisfaction of incumbents more broadly, consistent with the interpretation that mobile internet (and the social media platforms it enables) exposes citizens to critical political information that would otherwise be filtered out by pro-government broadcast media. The magnitudes are substantial, suggesting that the informational channel of social media operates not only within the confines of a thirty day deactivation study but also at the population level over longer time horizons.

6.2 Effects on Religious Tolerance and Organizational Participation

The second set of results I seek to corroborate with the instrumental variables analysis concerns the findings on religious tolerance. Again, I collate survey responses over the course of nearly three decades in which enumerators quizzed respondents about the extent to which they would be upset if members of another religion held a public event, built a house of worship, or became mayor in their community. I scale the responses such that higher values denote greater intolerance (i.e., responding “yes” to the statement). I code the values on a 0-1 scale such that the estimates can be interpreted as percentage point changes in the instrumented effect of internet usage on intolerance.

Table 2 reports the second-stage estimates for religious tolerance in columns 2-4. Column 1 reports the first stage estimates, again, showing that the exogenous instruments are strong negative predictors of internet usage on the part of survey respondents. Column 2 shows that a one-unit increase in predicted internet coverage reduces the probability of a respondent being upset by a member of religious organization holding an event in their neighborhood by 0.29 points on the response scale ($p < 0.01$). The decline increases in magnitude to a 0.35 point drop when asked if they would be upset if members of another religion built a church in their community ($p < 0.01$). Finally, column 4 shows that a one-unit increase in predicted internet coverage leads to a drop on the scale quizzing respondents about the extent to which they would be upset if they had a mayor from a different religion ($p < 0.01$).

Table 2: The Effect of Internet on Religion and Tolerance

	First stage	Second stage outcomes				
	Internet	Would you be upset if other religion...			Religious activities	
		Held event	Built church	Became mayor	Religious org	Study group
	(1)	(2)	(3)	(4)	(5)	(6)
Elevation SD (km)	-0.325*** (0.074)					
Dist. nearest city (log)	-0.118*** (0.013)					
Internet		-0.290*** (0.051)	-0.355*** (0.085)	-0.130*** (0.043)	-0.119*** (0.034)	-0.197* (0.115)
Observations	78,242	15,019	12,778	13,095	49,401	43,082
R-squared	0.126	-0.150	-0.262	0.026	-0.022	0.016
Year FE	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓
First-stage F	56.85	192.20	107.58	175.09	361.47	295.69

Notes: Two-stage least squares estimates of the effect of internet access on religious intolerance and organizational participation. Column (1) reports the first stage, regressing individual-level internet access on two terrain-based instruments: terrain ruggedness (standard deviation of elevation, in km) and the log distance to the nearest major city. Columns (2)–(4) report second-stage estimates for intolerance outcomes drawn from harmonized Indonesian surveys (1999–2024): (2) “Would you be upset if a member of another religion held a religious event in your community?”; (3) “Would you be upset if a member of another religion built a house of worship in your community?”; (4) “Would you be upset if a member of another religion became head of local government?” Higher values indicate greater intolerance. Columns (5)–(6) measure religious organizational activity: (5) participation in religious organizations (NU, Muhammadiyah, or other Islamic organizations); (6) frequency of attendance at religious study groups (*pengajian*). All specifications include year and province fixed effects. Standard errors clustered at the district level in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Columns 5 and 6 in Table 2 examine the substitution mechanism using the instrumental variables framework. Specifically, I draw on an identical and frequently-recurring survey item that asks whether a respondent is an active member, non-active member, or not a member of different religious organizations, such as Nahdlatul Ulama (NU), Muhammadiyah, or other more doctrinal Islamic organizations, such as DDII, FPI, or HTI. I also leverage a question in which respondents were asked about the frequency with which they participated in religious study groups (*pengajian*). The results presented in Table 2 provide direct corroboration for the substitution channel established in the experimental results. Instrumented internet access significantly reduces participation in religious organizations (-0.12 , $p < 0.01$) and religious study groups (-0.20 , $p < 0.10$). These estimates suggest that mobile internet displaces time previously allocated to religiously-tinged communal activities, which in the substitution framework are thought to be potential incubators of exclusionary attitudes. The instrumental variables evidence thus supports the pro-

posed causal chain: internet access pulls individuals away from religiously homogeneous offline settings, reducing exposure to religious identity cues and, in turn, lowering intolerance.

7 Robustness and Alternative Explanations

My preferred interpretation emphasizes two distinct channels. Improved perceptions of governance among deactivated users appear to reflect reduced access to salient information about incumbent performance, a claim for which I find mixed support in contemporaneous news knowledge. At the same time, deactivating social media seems to worsen religious chauvinism by shifting participants' time toward activities that may heighten intolerance, such as religious study groups and exposure to politically charged religious sermons. Still, alternative mechanisms could produce similar patterns. I therefore consider two such explanations and then assess robustness through two additional tests.

The first alternative centers on network diversity, particularly as it relates to the religious chauvinism outcome. Recent work argues that social media can expose users to more diverse social networks than they would encounter offline, challenging the “echo chambers” hypothesis (Asimovic et al. 2021; Asimovic, Nagler and Tucker 2023; Guess et al. 2018). Asimovic (2026), for example, shows how social media can provide accessible windows into outgroup worldviews that might otherwise remain misunderstood. Under this account, deactivation may have increased intolerance by reducing intergroup contact and thereby weakening the contact-based processes that inhibit outgroup hostility. I test this directly using endline survey items in which respondents identified the seven people with whom they interacted most frequently each week, along with each contact's role, ethnicity, religion, ideology, and time spent together. These data allow me to construct time-weighted diversity indices across ethnic, religious, ideological, and role dimensions, as pre-registered in H2a. Figure A10 shows little support for this explanation. Contrary to its predictions, individual-level deactivation increased the overall network diversity index by 0.08 SDs ($p < 0.10$, $q < 0.10$). Household deactivation produced no significant effect on the overall index, though it increased ethnic diversity by 0.04 SDs ($p < 0.10$, $q < 0.10$). The strongest household effect appears instead on role diversity, where respondents reported substantially less diversity in the *types* of people they interacted with each week.

The second alternative concerns social norms and second-order beliefs. Social media may create the impression that tolerance is widespread, either through exposure to pluralistic discourse or through platform curation of pro-social content. If deactivation removes this reference point, respondents may revert

to local community norms, which in many Indonesian settings may be more religiously conservative. On this view, the observed intolerance effects would reflect changes in perceived norms rather than shifts in behavior or organizational exposure. I test this possibility using a pre-registered second-order inaccuracy index (H2d), which measures how accurately respondents estimate the prevalence of various social groups in Indonesia, including Ahmadiyah adherents, Prabowo supporters, and Chinese-Indonesians. This is a demanding test because some minorities, especially the Ahmadiyah, who constitute less than 0.1% of the population, receive disproportionate attention in Indonesia (Soedirgo 2020). If social media shaped how respondents anchored beliefs about others' beliefs, deactivation should have shifted these estimates. Figure A10 provides no evidence of such an effect: the overall index and all seven constituent items are statistically indistinguishable from zero in both treatment arms.

I next examine robustness to model specification. Because the pre-registered specification did not include lagged dependent variables, I re-estimate the main models controlling for baseline standardized outcomes in Figures A6 and A7. The substantive conclusions remain largely unchanged. The main exception is the effect of household deactivation on the religious intolerance index: in the main specification this estimate was marginally significant under conventional p -values, though it did not survive multiple-testing correction, and with the baseline outcome included it becomes less precise ($p = 0.31$). This shift likely reflects slight pre-treatment imbalance across treatment arms on that outcome.

Finally, I probe the channel through which the effects operate. The control condition is “active” in that it entails continued social media use, which may drive the governance results, while the treatment conditions are also active in that they free time for alternative activities, which is my preferred explanation for the religious chauvinism results. Distinguishing these sources of movement is difficult when control-group values are normalized to zero. To address this, I plot pre- and post-treatment raw means for the individual survey items underlying the six outcome indices in Figures A8 and A9. The resulting patterns are consistent with the proposed channels: treatment appears to move the religious chauvinism outcomes, whereas continued exposure in the control condition appears to drive the governance-related outcomes.

8 Discussion

At least in the Indonesian context, the evidence presented here suggests that social media access increases exposure to critical political information, elevating perceptions of corruption in the short-run experimental evidence and reducing satisfaction with government in the longer-run instrumental variables analysis.

It also appears that social media displaces time spent in offline religious organizations, thereby reducing exposure to politicized religious content, and thus reducing religious intolerance and exclusionary nationalism. Removing access to social media evidently reverses both channels, as shown in the experimental and follow-up analyses. In what follows, I discuss two implications of these results for the broader literature on social media and politics. I conclude with a statement of some of the shortcomings of this research and outline what I view as potentially profitable directions for future work building on the findings contained in this paper.

The first implication of this research is to complicate the prevailing analytical framework in which social media's effects are understood primarily through users' direct exposure to online content. On this view, much of the scholarly and public debate has focused on the informational environment generated by social media platforms' feeds, thus emphasizing the potential of platforms to spread misinformation and fuel outrage via algorithmic curation of content (Bakshy, Messing and Adamic 2015; Brady et al. 2017; Vosoughi, Roy and Aral 2018). These mechanisms are surely important. But the evidence presented in this paper calls upon scholars to reflect on the appropriate counterfactual against which social media ought to be evaluated in a given context. The alternative to time spent on social media, after all, is not an informational vacuum.

I provide direct evidence consistent with this interpretation. The follow-up survey offers the clearest micro-level support: Muslim respondents assigned to the individual deactivation arm report greater participation in religious organizations during the study period. Existing scholarship on the Indonesian context has emphasized significant variation in the tolerance-inducing or -inhibiting effects of membership in different religious organizations in Indonesia; Menchik (2016), for example, argues that even if the major religious organizations in Indonesia (such as NU and Muhammadiyah) are not supporting Western-style liberalism they are nonetheless tolerance-promoting and may actually serve as institutions that support inter-religious harmony, possibly along the lines proposed by Fearon and Laitin (1996). The effects thus likely stem in part from increased participation in chauvinist institutions such as unguided religious study groups (*majelis taklim*) as well as more radical (and banned) Islamic organizations (e.g., FPI) whose rhetoric promotes more chauvinist worldviews.

The second implication of this research worth highlighting concerns the external validity of existing deactivation experiments. The dominant finding from the Western contexts in which such experiments have chiefly been conducted is that social media deactivation produces only modest effects on political

attitudes (Allcott et al. 2020, 2024; Larreguy and Raffler 2025). The findings contained in this paper suggest that this consensus may partly reflect the specific configuration of counterfactual social life in the United States and Western Europe. In settings marked by moribund associational life and weak social capital, the principal offline substitutes for social media are often comparatively individualized, which may limit their capacity to generate systematic attitudinal change. At the same time, alternative media sources, notably politicized cable news, may provide informational diets that are not substantively distinct from those encountered on social platforms, which may also reduce the scope for deactivation to shift political attitudes. In Indonesia, by contrast, the proximate substitutes when social media is deactivated are dense religious organizational networks. This suggests that the available substitutes to social media constitute an important scope condition—one that may vary across societies, but that prior experimental work has largely held constant by concentrating on a narrow set of Western cases.

The importance of considering substitutes to social media has large consequences for the growing number of governments that are considering introducing policies to restrict social media access. The results presented here suggest that such policies may produce unintended consequences if social media bans are introduced based on evidence generated in contexts with alternative social configurations. This possibility looms especially large in contexts where offline alternatives are less pluralistic than the online environment (a point also demonstrated by Chen and Yang 2019, in the context of media censorship in China). A government that shuts down social media to reduce the spread of “divisive content” may inadvertently channel citizens toward offline settings that foster more exclusionary attitudes than the platforms themselves (Enikolopov, Makarin and Petrova 2020; Campante, Durante and Sobbrío 2018; Manacorda and Tesei 2020).

By way of conclusion, I highlight several limitations that warrant discussion and ought to structure future research. For one, the experiment lasted thirty days, which was long enough to detect attitudinal shifts but potentially too short to capture equilibrium adjustments in media consumption and social behavior. Second, the experiment was conducted in Indonesia, and the specific substitution channels I identify, such as pro-government television and offline Islamic organizational networks, are features unique to this particular context. The general principle that social media’s effects depend on what it displaces should travel broadly, but the specific direction and magnitude of effects will vary with the underlying local media ecosystem and civil society landscape—a possibility that future research should take seriously by undertaking more studies that consider the effects of social media in comparative perspective.

References

- Allcott, Hunt, Luca Braghieri, Sarah Enikopolov and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review* 110(3):629–676.
- Allcott, Hunt, Luca Braghieri, Sarah Enikopolov, Matthew Gentzkow and Chuan Pei. 2024. "The Effects of Facebook and Instagram on the 2020 Election: A Deactivation Experiment." *Proceedings of the National Academy of Sciences* 121(21).
- Allen, Jennifer and Joshua A Tucker. 2025. "Platform-independent experiments on social media." *Science* 390(6776):883–884.
- Altay, Sacha, Emma Hoes and Magdalena Wojcieszak. 2025. "Following News on Social Media Boosts Knowledge, Belief Accuracy and Trust." *Nature Human Behaviour* 9:1833–1842.
- Amar, Priyadarshi, Sumitra Badrinathan, Simon Chauchard and Florian Sichart. 2025. "Countering misinformation early: Evidence from a classroom-based field experiment in India." *American Political Science Review* pp. 1–21.
- Arceneaux, Kevin, Can Zengin and Jonathan M Ladd. 2025. "Stop Blaming Social Media for Everything: The Minimal Effects of Facebook in the 2024 US Presidential Election."
- Arceneaux, Kevin, Martial Foucault, Kalli Giannelos, Jonathan Ladd and Can Zengin. 2024. "Facebook increases political knowledge, reduces well-being and informational treatments do little to help." *Royal Society Open Science* 11(10):240280.
- Asimovic, Nejla. 2026. "Unlocking Outgroup Access Online: Evidence From Cyprus." *Comparative Political Studies*.
- Asimovic, Nejla, Jonathan Nagler and Joshua A Tucker. 2023. "Replicating the effects of Facebook deactivation in an ethnically polarized setting." *Research & Politics* 10(4):20531680231205157.
- Asimovic, Nejla, Jonathan Nagler, Richard Bonneau and Joshua A. Tucker. 2021. "Testing the Effects of Facebook Usage in an Ethnically Polarized Setting." *Proceedings of the National Academy of Sciences* 118(25):e2022819118.
- Aspinall, Edward and Marcus Mietzner. 2019. "Southeast Asia's troubling elections: Nondemocratic pluralism in Indonesia." *Journal of Democracy* 30(4):104–118.
- Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout and Alexander Volfovsky. 2018. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." *Proceedings of the National Academy of Sciences* 115(37):9216–9221.
- Bakshy, Eytan, Solomon Messing and Lada A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348(6239):1130–1132.
- Barron, Patrick, Sana Jaffrey and Ashutosh Varshney. 2016. "When Large Conflicts Subside: The Ebbs and Flows of Violence in Post-Suharto Indonesia." *Journal of East Asian Studies* 16(2):191–217.
- Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)* 57(1):289–300.

- Bowers, Jake, Mark M Fredrickson and Costas Panagopoulos. 2013. "Reasoning about interference between units: A general framework." *Political Analysis* 21(1):97–124.
- Bowles, Jeremy, John Marshall and Pia Raffler. 2024. "Access to Social Media and Support for Elected Autocrats: Field Experimental and Observational Evidence from Uganda." Unpublished manuscript.
- Brady, William J., Julian A. Wills, John T. Jost, Joshua A. Tucker and Jay J. Van Bavel. 2017. "Emotion Shapes the Diffusion of Moralized Content in Social Networks." *Proceedings of the National Academy of Sciences* 114(28):7313–7318.
- Bursztyn, Leonardo, Benjamin Handel, Rafael Jiménez-Durán and Christopher Roth. 2025. "When product markets become collective traps: The case of social media." *American Economic Review* 115(12):4105–4136.
- Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov and Maria Petrova. 2019. Social media and xenophobia: evidence from Russia. Technical report National Bureau of Economic Research.
- Campante, Filipe R., Ruben Durante and Andrea Tesei. 2022. "Media and Social Capital." *Annual Review of Economics* 14:69–91.
- Campante, Filipe, Ruben Durante and Francesco Sobbrío. 2018. "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.
- Djourelouva, Milena, Ruben Durante and Gregory J Martin. 2025. "The impact of online competition on local newspapers: Evidence from the introduction of Craigslist." *Review of Economic Studies* 92(3):1738–1772.
- Donati, Dante. 2023. "Mobile Internet access and political outcomes: Evidence from South Africa." *Journal of Development Economics* 162:103073.
- Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D Hyde, Craig McIntosh and Gareth Nellis. 2019. *Information, accountability, and cumulative learning: Lessons from Metaketa I*. Cambridge University Press.
- Egami, Naoki. 2021. "Spillover effects in the presence of unobserved networks." *Political Analysis* 29(3):287–316.
- Enikolopov, Ruben, Alexey Makarin and Maria Petrova. 2020. "Social Media and Protest Participation: Evidence from Russia." *Econometrica* 88(4):1479–1514.
- Fearon, James D and David D Laitin. 1996. "Explaining interethnic cooperation." *American political science review* 90(4):715–735.
- Gentzkow, Matthew. 2006. "Television and Voter Turnout." *The Quarterly Journal of Economics* 121(3):931–972.
- Guess, Andrew, Brendan Nyhan, Benjamin Lyons and Jason Reifler. 2018. "Avoiding the echo chamber about echo chambers." *Knight Foundation* 2(1):1–25.
- Guess, Andrew, Jonathan Nagler and Joshua Tucker. 2019. "Less than you think: Prevalence and predictors of fake news dissemination on Facebook." *Science advances* 5(1):eaau4586.

- Guess, Andrew M, Brendan Nyhan and Jason Reifler. 2020. "Exposure to untrustworthy websites in the 2016 US election." *Nature human behaviour* 4(5):472–480.
- Guess, Andrew M., Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow et al. 2023. "How Do Social Media Feed Algorithms Affect Attitudes and Behavior in an Election Campaign?" *Science* 381(6656):398–404.
- Guriey, Sergei, Nikita Melnikov and Ekaterina Zhuravskaya. 2021. "3G Internet and Confidence in Government." *The Quarterly Journal of Economics* 136(4):2533–2613.
- Hopkins, Daniel J. 2018. *The increasingly United States: How and why American political behavior nationalized*. University of Chicago Press.
- Jaffrey, Sana. 2023. "Special symposium, collective vigilantism in global comparative perspective mechanics of impunity: Vigilantism and state-building in Indonesia." *Comparative Politics* 55(2):287–311.
- Kiratli, Osman Sabri. 2023. "Social Media Effects on Public Trust in the European Union." *Public Opinion Quarterly* 87(3):749–763.
- Klein, Elad and Joshua Robison. 2020. "Like, Post, and Distrust? How Social Media Use Affects Trust in Government." *Political Communication* 37(1):46–64.
- Kling, Jeffrey R., Jeffrey B. Liebman and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1):83–119.
- Kraut, Robert, Michael Patterson, Vicki Lundmark, Sara Kiesler, Tridas Mukophadhyay and William Scherlis. 1998. "Internet Paradox: A Social Technology That Reduces Social Involvement and Psychological Well-Being?" *American Psychologist* 53(9):1017–1031.
- Kuipers, Nicholas. 2026. "Pluralism, Polarization, and Political Voyeurism." *Journal of Democracy* .
- Kuipers, Nicholas, Deni Irvani and Saiful Mujani. n.d. "The Mental Health Benefits of Social Media Deactivation under Collective Participation." <https://osf.io/gu5pq/files/ykfsu>. Unpublished manuscript.
- Kuipers, Nicholas, Gareth Nellis and Michael Weaver. 2021. "Does Electing Extremist Parties Increase Violence and Intolerance?" *British Journal of Political Science* 51(3):1340–1347.
- Larreguy, Horacio and Pia J Raffler. 2025. "Accountability in developing democracies: the impact of the internet, social media, and polarization." *Annual Review of Political Science* 28.
- Levy, Ro'ee. 2021. "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." *American Economic Review* 111(3):831–870.
- Manacorda, Marco and Andrea Tesei. 2020. "Liberation Technology: Mobile Phones and Political Mobilization in Africa." *Econometrica* 88(2):533–567.
- McCarty, Nolan. 2019. *Polarization: What everyone needs to know®*. Oxford University Press.
- McKenzie, David. 2012. "Beyond Baseline and Follow-Up: The Case for More T in Experiments." *Journal of Development Economics* 99(2):210–221.
- Menchik, Jeremy. 2016. *Islam and democracy in Indonesia: Tolerance without liberalism*. Cambridge University Press.

- Mietzner, Marcus. 2026. *Ruling Indonesia: Jokowi's Presidency in an Age of Democratic Crisis and Great Power Competition*. University of Michigan Press.
- Mietzner, Marcus, Burhanuddin Muhtadi and Rizka Halida. 2018. "Entrepreneurs of grievance: Drivers and effects of Indonesia's Islamist mobilization." *Bijdragen tot de taal-, land-en volkenkunde/Journal of the Humanities and Social Sciences of Southeast Asia* 174(2-3):159–187.
- Moskowitz, Daniel J. 2021. "Local news, information, and the nationalization of US elections." *American Political Science Review* 115(1):114–129.
- Mujani, Saiful and Nicholas Kuipers. 2020. "Who believed misinformation during the 2019 Indonesian election?" *Asian Survey* 60(6):1029–1043.
- Mujani, Saiful and R. William Liddle. 2009. "Muslim Indonesia's Secular Democracy." *Asian Survey* 49(4):575–590.
- Müller, Karsten and Carlo Schwarz. 2021. "Fanning the Flames of Hate: Social Media and Hate Crime." *Journal of the European Economic Association* 19(4):2131–2167.
- Müller, Karsten and Carlo Schwarz. 2023. "From Hashtag to Hate Crime: Twitter and Antiminority Sentiment." *American Economic Journal: Applied Economics* 15(3):270–312.
- Nyhan, Brendan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y. Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery et al. 2023. "Like-Minded Sources on Facebook Are Prevalent but Not Polarizing." *Nature* 620:137–144.
- Olken, Benjamin A. 2009. "Do Television and Radio Destroy Social Capital? Evidence from Indonesian Villages." *American Economic Journal: Applied Economics* 1(4):1–33.
- Pepinsky, Thomas B, R William Liddle and Saiful Mujani. 2018. *Piety and Public Opinion: Understanding Indonesian Islam*. Oxford University Press.
- Piccardi, Tiziano, Martin Saveski, Chenyan Jia, Jeffrey Hancock, Jeanne L Tsai and Michael S Bernstein. 2025. "Reranking partisan animosity in algorithmic social media feeds alters affective polarization." *Science* 390(6776):eadu5584.
- Prior, Markus. 2007. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge University Press.
- Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.
- Samet, Oren, Leonardo R Arriola and Aila M Matanock. 2024. "Facebook usage and outgroup intolerance in Myanmar." *Political Communication* 41(6):944–964.
- Setijadi, Charlotte. 2017. *Ahok's downfall and the rise of Islamist populism in Indonesia*. ISEAS-Yusof Ishak Institute Singapore.
- Soderborg, Seth and Burhanuddin Muhtadi. 2023. "Resentment and polarization in Indonesia." *Journal of East Asian Studies* 23(3):439–467.
- Soedirgo, Jessica. 2020. *The threat of small things: Patterns of repression and mobilization against micro-sized groups in Indonesia*. University of Toronto (Canada).

- Stevic, Anja, Desirée Schmuck, Kathrin Karsay and Jörg Matthes. 2021. "Are Smartphones Enhancing or Displacing Face-to-Face Communication With Close Ties? A Panel Study Among Adults." *International Journal of Communication* 15:792–813.
- Sunstein, Cass R. 2018. *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.
- Tapsell, Ross. 2017. *Media power in Indonesia: Oligarchs, citizens and the digital revolution*. Bloomsbury Publishing USA.
- Torreblanca, Carolina. 2025. "The Demobilizing Effect of Violence: Evidence of Heterogeneous Exposure and Response." Working paper.
URL: <https://carolina-torreblanca.github.io/files/papers/crime/paper.pdf>
- Tucker, Joshua A., Yannis Theocharis, Margaret E. Roberts and Pablo Barberá. 2017. "From Liberation to Turmoil: Social Media and Democracy." *Journal of Democracy* 28(4):46–59.
- Twenge, Jean M, Jonathan Haidt, Thomas E Joiner and W Keith Campbell. 2020. "Underestimating digital media harm." *Nature Human Behaviour* 4(4):346–348.
- Ventura, Tiago, Rajeshwari Majumdar, Jonathan Nagler and Joshua Tucker. Forthcoming. "Misinformation Beyond Traditional Feeds: Evidence from a WhatsApp Deactivation Experiment in Brazil." *Journal of Politics*.
- Vosoughi, Soroush, Deb Roy and Sinan Aral. 2018. "The spread of true and false news online." *science* 359(6380):1146–1151.
- Warburton, Eve. 2020. "Deepening polarization and democratic decline in Indonesia." *Political Polarization in South and Southeast Asia* 17.

Supplemental Appendix

A Preliminaries	2
A.1 Ethical Considerations	2
A.2 CONSORT Flow Diagram	3
A.3 Pre-analysis Plan Deviations	4
A.4 Platform-Level Compliance	5
A.5 Experimental Survey Questions	6
A.6 IV Data Construction Description	8
B Descriptive Statistics	9
B.1 Experimental Design	9
B.2 Instrumental Variables Design	14
C Additional Analyses	15
C.1 Experimental Design	15
C.1.1 Local Average Treatment Effect (LATE) Estimation	15
C.1.2 Specification with Pre-Treatment Controls	18
C.1.3 Raw Outcome Means by Treatment Group	20
C.1.4 Tabular Regression Estimates	22
C.1.5 Alternative Explanations	25
C.2 Instrumental Variables Design	26
C.2.1 Robustness: Distance-to-City Controls	26
C.2.2 Robustness: Village-Level Economic Development Controls	27

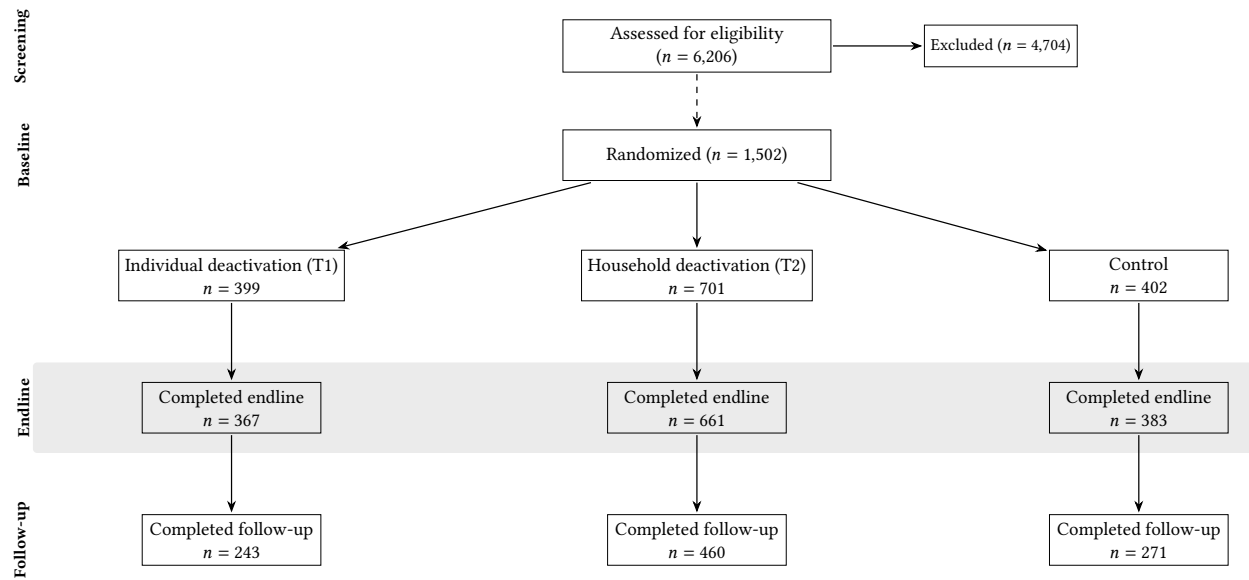
A Preliminaries

A.1 Ethical Considerations

All research protocols were approved by [REDACTED]. The local implementing partner, Saiful Mujani Research and Consulting (SMRC), maintains its own internal review process and issued its own approval. No deception was involved in the study: potential participants were informed about all three experimental conditions and the goal of the research at the outset. Financial incentives (400,000–600,000 IDR, approximately \$25–38 USD) were calibrated to compensate for potential lost utility from withdrawing from social media for four weeks. We limited our study population to those that did not use social media for business reasons, both to minimize the potential for adverse financial harm and to ensure that incentives compensated for foregone leisure rather than lost income. Participants could withdraw from the study at any time and still receive their promised incentive at no cost. The monitoring application (SosMedMonitoring) tracked only aggregate time-on-app data, not browsing content or the substance of participants’ online activity, thereby limiting exposure to adverse consequences in the event of a data breach.

A.2 CONSORT Flow Diagram

Figure A1—CONSORT flow diagram.



Note: Dashed arrow indicates the point of randomization. Participants were randomized into three arms: individual social media deactivation (T1), household-level deactivation (T2), and control. The endline survey was administered four weeks after baseline. The follow-up survey was administered after the deactivation period ended. Attrition from baseline to endline was 6.1%; from baseline to follow-up was 35.2%.

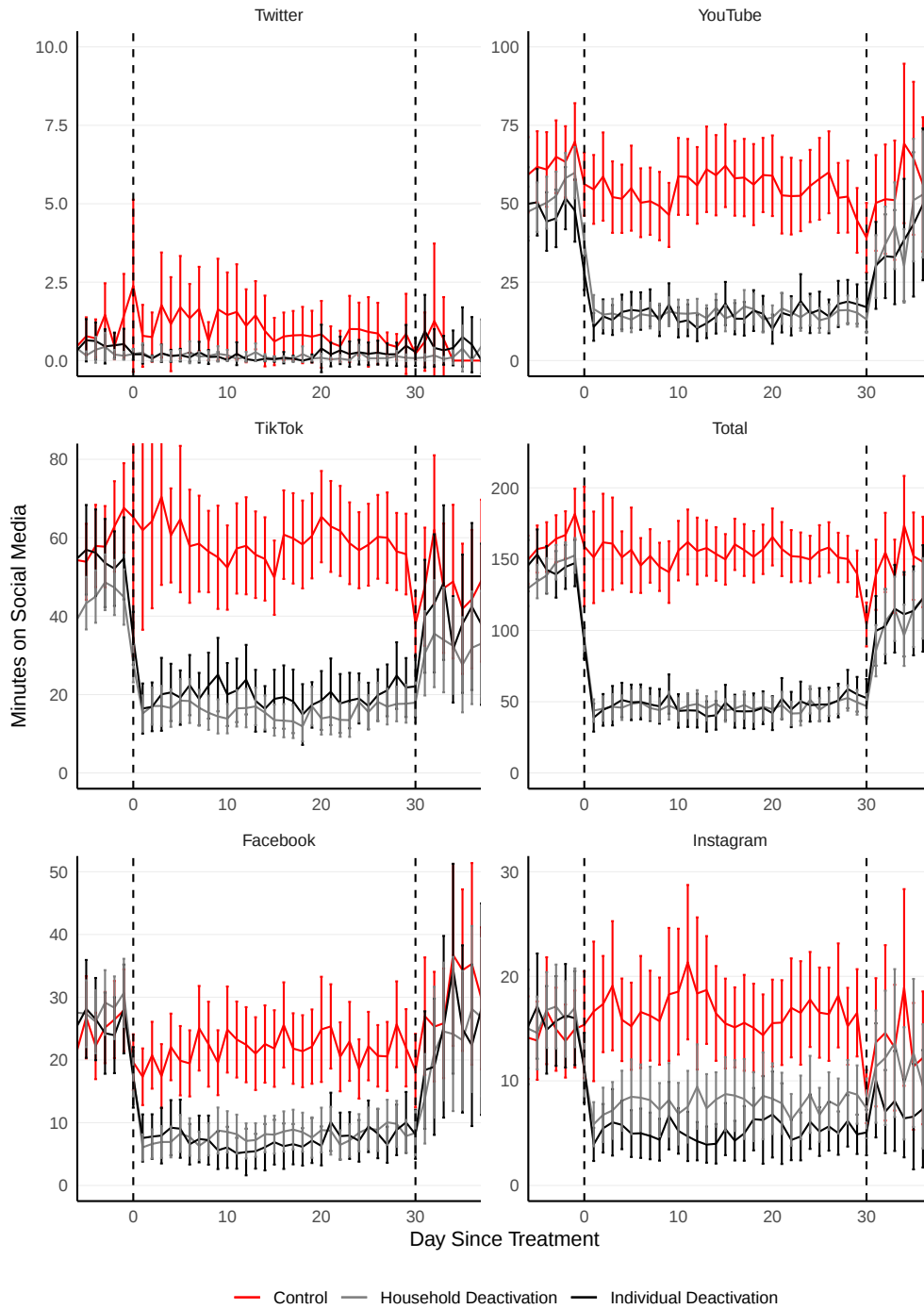
A.3 Pre-analysis Plan Deviations

The key elements of the research design and outcome definitions were pre-registered prior to data collection. The following deviations from the pre-analysis plan are noted:

- **Outcome reorganization and addition.** The PAP organized outcomes under hypothesis-specific families (H1a: well-being; H1b: political knowledge; H1c: outgroup animosity; H1d: government satisfaction; H2a–e: secondary outcomes). The manuscript reorganizes these into thematically coherent families—corruption perceptions, government support, information accuracy, religious intolerance, support for religious violence, and national identity—which better map onto the substantive interpretation of results. Several of the outcomes and indices were not pre-registered as part of the main indices (e.g., corruption perceptions, support for religious violence, religious nationalism).
- **Post-hoc substitution framework.** The theoretical framework in Section 2, which interprets the experimental results through the lens of time-use substitution toward television and offline religious participation, was developed after the experimental results were known. It was not pre-registered. I flag this framework as explicitly post-hoc and use complementary evidence (follow-up survey, instrumental variables analysis) to assess its plausibility.
- **Follow-up mechanism survey.** The follow-up survey measuring television substitution and religious organization participation was designed after the main experimental results were known to test the substitution framework described above. It was not pre-registered.
- **Unreported outcomes.** Several PAP secondary outcomes—economic productivity (H2b), social comparison (H2c), and political voyeurism (H2e)—are not reported in the main figures but are available upon request. All except that last (H2e) were null results. One of the primary PAP outcomes (mental health and subjective wellbeing) was written up as part of a separate paper and is hence not included in the present analysis. It is referenced in the main text. The deactivation intervention yielded large improvements in self-reported mental health.

A.4 Platform-Level Compliance

Figure A2—Platform-level compliance time series.



Note: Daily social media usage (minutes) by platform and treatment arm over the experimental period.

A.5 Experimental Survey Questions

Table A1: Survey question wording and index membership.

Variable	Question wording	Scale
<i>Corruption perceptions</i>		
V_21	Menurut Ibu/Bapak, seberapa sering pejabat pemerintah menyembunyikan informasi penting dari masyarakat?	1–4
V_28	Menurut Ibu/Bapak, bagaimana praktik korupsi dan suap di kalangan pejabat pemerintah saat ini? Apakah sangat rendah, rendah, tinggi, atau sangat tinggi?	1–4
<i>Government support</i>		
V_13	Secara umum, apakah sejauh ini Ibu/Bapak sangat puas, cukup puas, kurang puas, atau tidak puas sama sekali dengan kerja Presiden Prabowo Subianto dalam menjalankan pemerintahan?	1–4
V_27_A	Seberapa besar kepercayaan Ibu/Bapak terhadap lembaga-lembaga berikut ini? — Presiden	1–4
V_27_B	Dewan Perwakilan Rakyat (DPR)	1–4
V_27_C	Pemerintah Daerah (kabupaten/kota)	1–4
<i>Information accuracy</i>		
V_51_A	Menteri keuangan Indonesia sekarang adalah Sri Mulyani	T/F
V_51_B	Kepala Kepolisian Republik Indonesia (Kapolri) Jenderal Listyo Sigit Prabowo telah diganti	T/F
V_51_C	Pemerintah Provinsi Jakarta berencana memindahkan kampus Institut Kesenian Jakarta (IKJ) pindah ke Lebak Bulus	T/F
V_51_D	Pemerintah Provinsi Jakarta berupaya mengatasi krisis sampah dengan mengoperasikan RDF Plant (Refuse Derived Fuel) Rorotan	T/F
V_51_E	Sekitar 40% Dana Pendidikan dari Anggaran Pendapatan dan Belanja Negara (APBN) dialokasikan untuk Makan Bergizi Gratis (MBG)	T/F
V_51_F	Gubernur Maluku Utara Sherly Tjoanda adalah istri mendiang mantan Bupati Pulau Morotai	T/F
<i>Non-Muslim intolerance</i>		
V_31_A	Apakah Ibu/Bapak keberatan jika: — Warga non-Muslim melakukan kegiatan keagamaan di lingkungan tempat tinggal Ibu/Bapak.	1–4
V_31_B	Warga non-Muslim mendirikan rumah ibadah di lingkungan tempat tinggal Ibu/Bapak.	1–4
V_31_C	Non-Muslim menjadi kepala daerah di wilayah mayoritas Muslim.	1–4
<i>Political violence</i>		
V_30_A	Sejauh mana Ibu/Bapak menganggap kekerasan dapat dibenarkan untuk: — Untuk mencegah kecurangan dalam pemilu	1–4
V_30_B	Untuk melawan pengaruh negatif yang merusak moral masyarakat	1–4
V_30_C	Untuk melindungi kehormatan agama Ibu/Bapak dari penistaan	1–4
<i>National identity</i>		
V_18	Secara umum, seberapa bangga Ibu/Bapak menjadi warga negara Indonesia?	1–4
V_20	Menurut Ibu/Bapak, apakah Pancasila masih relevan/cocok sebagai dasar panduan bagi bangsa Indonesia di era modern ini?	1–4
<i>Media use</i>		
V_21A	Dengan mempertimbangkan empat minggu terakhir secara khusus, berapa banyak waktu rata-rata per hari yang Anda habiskan untuk media sosial?	1–5
V_22_A	Dalam empat minggu terakhir, seberapa sering Ibu/Bapak menggunakan media berikut untuk mendapatkan berita atau informasi politik? — Media cetak	1–4
V_22_B	Portal berita online	1–4
V_22_C	Instagram	1–4
V_22_D	Televisi	1–4
V_22_E	Facebook	1–4
V_22_F	Twitter	1–4
V_22_G	TikTok	1–4
<i>Demographics</i>		
V_1	Jenis Kelamin Ibu/Bapak	1–2
V_2	[HARUS BERUSIA 18 TAHUN KE ATAS] Berapa umur Ibu/Bapak sekarang?	years
V_3	Apa pendidikan terakhir Ibu/Bapak?	1–7
V_5	Rata-rata dalam sebulan, kira-kira berapa total pendapatan kotor rumah tangga Ibu/Bapak? Rumah tangga berarti semua orang yang tinggal bersama Ibu/Bapak dalam satu rumah dan pengelolaan makannya dari satu dapur.	1–7
V_6	Apa agama atau kepercayaan Ibu/Bapak?	1–8
V_7	Selain sebagai warga negara Indonesia, Ibu/Bapak termasuk dalam kelompok etnis apa?	1–11

Table A2: Survey question wording and index membership (English).

Variable	Question wording	Scale
<u>Corruption perceptions</u>		
V_21	In your opinion, how often do government officials hide important information from the public?	1–4
V_28	In your opinion, what is the current level of corruption and bribery among government officials? Is it very low, low, high, or very high?	1–4
<u>Government support</u>		
V_13	Overall, so far are you very satisfied, somewhat satisfied, somewhat dissatisfied, or very dissatisfied with President Prabowo Subianto's performance in running the government?	1–4
V_27_A	How much do you trust the following institutions? — President	1–4
V_27_B	House of Representatives (DPR)	1–4
V_27_C	Local government (regency/city)	1–4
<u>Information accuracy</u>		
V_51_A	Indonesia's current Minister of Finance is Sri Mulyani	T/F
V_51_B	National Police Chief (Kapolri) General Listyo Sigit Prabowo has been replaced	T/F
V_51_C	The Jakarta Provincial Government plans to relocate the Jakarta Institute of the Arts (IKJ) campus to Lebak Bulus	T/F
V_51_D	The Jakarta Provincial Government is trying to address the waste crisis by operating the Rorotan RDF Plant (Refuse Derived Fuel)	T/F
V_51_E	Around 40% of the Education Budget from the State Budget (APBN) is allocated to the Free Nutritious Meals (MBG) program	T/F
V_51_F	North Maluku Governor Sherly Tjoanda is the wife of the late former Regent of Morotai Island	T/F
<u>Non-Muslim intolerance</u>		
V_31_A	Would you be upset if: — Non-Muslim residents carry out religious activities in your neighborhood.	1–4
V_31_B	Non-Muslim residents establish a house of worship in your neighborhood.	1–4
V_31_C	A non-Muslim becomes a regional head in a Muslim-majority area.	1–4
<u>Political violence</u>		
V_30_A	To what extent do you think violence is justified to: — To prevent fraud in elections	1–4
V_30_B	To fight negative influences that damage public morality	1–4
V_30_C	To protect the honor of your religion from blasphemy	1–4
<u>National identity</u>		
V_18	Overall, how proud are you to be an Indonesian citizen?	1–4
V_20	In your opinion, is Pancasila still relevant/suitable as a foundational guide for the Indonesian nation in this modern era?	1–4
<u>Media use</u>		
V_21A	Thinking specifically about the last four weeks, on average how much time per day have you spent on social media?	1–5
V_22_A	In the last four weeks, how often did you use the following media to get news or political information? — Print media	1–4
V_22_B	Online news portals	1–4
V_22_C	Instagram	1–4
V_22_D	Television	1–4
V_22_E	Facebook	1–4
V_22_F	Twitter	1–4
V_22_G	TikTok	1–4
<u>Demographics</u>		
V_1	Your gender	1–2
V_2	[MUST BE 18 YEARS OLD OR ABOVE] How old are you now?	years
V_3	What is your highest level of education?	1–7
V_5	On average in a month, approximately what is your household's total gross income? Household means everyone who lives with you in one house and whose meals are managed from one kitchen.	1–7
V_6	What is your religion or belief?	1–8
V_7	Other than being an Indonesian citizen, which ethnic group do you belong to?	1–11

A.6 IV Data Construction Description

The instrumental variables analysis draws on three data sources, which I describe in turn.

Harmonized Survey Outcomes I harmonize individual-level survey data from three major Indonesian survey companies spanning 1999–2024. In total, I draw on 152 survey waves across these three survey houses. I standardize variable names and response scales across waves to construct a pooled panel of individual-survey-year observations. Key outcome variables include satisfaction with the incumbent president (1–4 scale), perceptions of corruption among government officials (1–4), satisfaction with democracy (1–4), religious intolerance items (binary indicators for whether the respondent would be upset if a member of another religion held a public event, built a house of worship, or became head of local government), participation in religious organizations (membership in NU, Muhammadiyah, or other Islamic organizations), and frequency of attendance at religious study groups (*pengajian*).

Terrain Instruments The instrumental variables strategy exploits plausibly exogenous variation in the cost of mobile network infrastructure deployment driven by terrain characteristics. Following Olken (2009), I construct two instruments at the village level: (1) the standard deviation of elevation within the village boundary (terrain ruggedness), which proxies for the difficulty of installing cell towers in mountainous terrain; and (2) the log distance from the village centroid to the nearest major city, which captures the cost of extending network infrastructure to remote locations. Both instruments are strong negative predictors of mobile internet coverage, and I argue that conditional on province fixed effects and year fixed effects, they satisfy the exclusion restriction because these geographic features affect political attitudes only through their effect on communications infrastructure.

Village-Level Economic Controls To address the concern that terrain characteristics may correlate with village-level economic development, I draw two additional control variables from the 2014 Village Census (PODES, *Potensi Desa*), a comprehensive census of all Indonesian villages conducted by BPS. First, I construct the share of households with electricity (PLN or non-PLN) as a proxy for local economic infrastructure and modernization (PODES variables R501A1, R501A2, R501B). Second, I extract the reported travel time in hours from the village to the district capital office (*kantor bupati/walikota*; PODES variable R1002B_K3), which captures administrative remoteness and ease of access to government services. These variables are matched to the survey data via 10-digit BPS village codes. I use them in robustness checks (Tables A13 and A14) to verify that the IV estimates are not driven by omitted village-level economic conditions.

B Descriptive Statistics

B.1 Experimental Design

Table A3: Summary statistics (Demographics: baseline vs endline samples).

Measure	N	Mean	St. Dev.	Min.	Median	Max.
<u>Baseline</u>						
Age (baseline)	1502	41.81	13.36	18	43	92
Gender: male (share)	1502	0.50	0.50	0	0	1
Gender: female (share)	1502	0.50	0.50	0	0	1
Gender: other/unknown (share)	1502	0.00	0.06	0	0	1
Education: primary or less (share)	1502	0.38	0.49	0	0	1
Education: secondary (share)	1502	0.40	0.49	0	0	1
Education: tertiary+ (share)	1502	0.18	0.38	0	0	1
Education: other/unknown (share)	1502	0.04	0.19	0	0	1
Religion: Muslim (share)	1502	0.91	0.29	0	1	1
Religion: non-Muslim (share)	1502	0.09	0.28	0	0	1
Religion: other/unknown (share)	1502	0.01	0.07	0	0	1
Ethnicity: Javanese (share)	1502	0.02	0.14	0	0	1
Ethnicity: other (share)	1502	0.98	0.14	0	1	1
Assigned: control (share)	1502	0.27	0.44	0	0	1
Assigned: treatment 1 (share)	1502	0.27	0.44	0	0	1
Assigned: treatment 2 (share)	1502	0.47	0.50	0	0	1
<u>Endline</u>						
Age (baseline)	1406	41.79	13.13	18	43	92
Gender: male (share)	1406	0.49	0.50	0	0	1
Gender: female (share)	1406	0.50	0.50	0	1	1
Gender: other/unknown (share)	1406	0.00	0.07	0	0	1
Education: primary or less (share)	1406	0.38	0.49	0	0	1
Education: secondary (share)	1406	0.40	0.49	0	0	1
Education: tertiary+ (share)	1406	0.18	0.38	0	0	1
Education: other/unknown (share)	1406	0.03	0.18	0	0	1
Religion: Muslim (share)	1406	0.91	0.29	0	1	1
Religion: non-Muslim (share)	1406	0.09	0.28	0	0	1
Religion: other/unknown (share)	1406	0.00	0.07	0	0	1
Ethnicity: Javanese (share)	1406	0.02	0.14	0	0	1
Ethnicity: other (share)	1406	0.98	0.14	0	1	1
Assigned: control (share)	1406	0.27	0.44	0	0	1
Assigned: treatment 1 (share)	1406	0.26	0.44	0	0	1
Assigned: treatment 2 (share)	1406	0.47	0.50	0	0	1

Table A4: Summary statistics (raw index sub-items: baseline vs endline).

Measure	N	Mean	St. Dev.	Min.	Median	Max.
<i>Baseline (y0_raw)</i>						
Gov support index: Pres. job approval	1473	3.01	0.59	1	3	4
Gov support index: Trust in President	1471	3.15	0.55	1	3	4
Gov support index: Trust in parliament	1447	2.54	0.76	1	3	4
Gov support index: Trust in local govt	1446	2.90	0.59	1	3	4
Gov support index: Low corruption (rev.)	1429	1.87	0.76	1	2	4
Corruption perceptions: Officials hide info	1391	2.61	0.79	1	3	4
Corruption perceptions: Corruption level	1429	3.13	0.76	1	3	4
Political violence: Prevent election fraud	1432	1.94	1.03	1	2	4
Political violence: Fight moral decay	1431	2.08	1.04	1	2	4
Political violence: Defend religion honor	1437	2.24	1.11	1	2	4
Non-Muslim intolerance: Local worship activity	1325	2.14	0.70	1	2	4
Non-Muslim intolerance: Build house of worship	1320	2.27	0.75	1	2	4
Non-Muslim intolerance: Non-Muslim local leader	1314	2.41	0.81	1	2	4
Info accuracy: Finance minister is Sri Mulyani	1263	0.75	0.43	0	1	1
Info accuracy: Kapolri replaced (claim)	970	0.49	0.50	0	0	1
Info accuracy: IKJ campus moved (claim)	645	0.47	0.50	0	0	1
Info accuracy: RDF Rorotan operated (claim)	791	0.77	0.42	0	1	1
Info accuracy: 40 edu budget to MBG (claim)	1019	0.18	0.38	0	0	1
Info accuracy: North Maluku governor wife (claim)	791	0.77	0.42	0	1	1
<i>Endline (y1_raw)</i>						
Gov support index: Pres. job approval	1382	2.92	0.64	1	3	4
Gov support index: Trust in President	1346	3.14	0.63	1	3	4
Gov support index: Trust in parliament	1319	2.48	0.82	1	3	4
Gov support index: Trust in local govt	1324	2.85	0.63	1	3	4
Gov support index: Low corruption (rev.)	1314	1.83	0.77	1	2	4
Corruption perceptions: Officials hide info	1261	2.70	0.79	1	3	4
Corruption perceptions: Corruption level	1314	3.17	0.77	1	3	4
Political violence: Prevent election fraud	1305	1.95	1.04	1	2	4
Political violence: Fight moral decay	1310	2.20	1.07	1	2	4
Political violence: Defend religion honor	1313	2.32	1.13	1	2	4
Non-Muslim intolerance: Local worship activity	1223	2.09	0.68	1	2	4
Non-Muslim intolerance: Build house of worship	1207	2.22	0.75	1	2	4
Non-Muslim intolerance: Non-Muslim local leader	1202	2.40	0.81	1	2	4
Info accuracy: Finance minister is Sri Mulyani	1186	0.77	0.42	0	1	1
Info accuracy: Kapolri replaced (claim)	913	0.54	0.50	0	1	1
Info accuracy: IKJ campus moved (claim)	604	0.50	0.50	0	1	1
Info accuracy: RDF Rorotan operated (claim)	733	0.80	0.40	0	1	1
Info accuracy: 40 edu budget to MBG (claim)	943	0.17	0.38	0	0	1
Info accuracy: North Maluku governor wife (claim)	745	0.81	0.39	0	1	1

Table A5: Balance of demographic traits across treatment arms (baseline and endline samples).

Variable	Control	Individual (T1)	Household (T2)	p-value
<i>Baseline</i>				
Age (baseline)	41.39 (13.37)	42.52 (13.46)	41.64 (13.31)	0.442
Gender: male (share)	0.51 (0.50)	0.49 (0.50)	0.49 (0.50)	0.693
Gender: female (share)	0.49 (0.50)	0.50 (0.50)	0.50 (0.50)	0.800
Gender: other/unknown (share)	0.00 (0.00)	0.01 (0.07)	0.01 (0.08)	0.328
Education: primary or less (share)	0.38 (0.48)	0.39 (0.49)	0.38 (0.49)	0.949
Education: secondary (share)	0.40 (0.49)	0.40 (0.49)	0.41 (0.49)	0.908
Education: tertiary+ (share)	0.17 (0.38)	0.19 (0.39)	0.17 (0.38)	0.857
Education: other/unknown (share)	0.05 (0.21)	0.03 (0.18)	0.03 (0.18)	0.463
Religion: Muslim (share)	0.90 (0.31)	0.91 (0.29)	0.91 (0.28)	0.659
Religion: non-Muslim (share)	0.09 (0.29)	0.09 (0.28)	0.08 (0.28)	0.843
Religion: other/unknown (share)	0.01 (0.10)	0.00 (0.05)	0.00 (0.07)	0.307
Ethnicity: Javanese (share)	0.02 (0.13)	0.02 (0.15)	0.02 (0.14)	0.861
Ethnicity: other (share)	0.98 (0.13)	0.98 (0.15)	0.98 (0.14)	0.861
<i>Endline</i>				
Age (baseline)	41.55 (13.36)	42.30 (12.99)	41.64 (13.09)	0.686
Gender: male (share)	0.52 (0.50)	0.49 (0.50)	0.48 (0.50)	0.501
Gender: female (share)	0.48 (0.50)	0.51 (0.50)	0.51 (0.50)	0.618
Gender: other/unknown (share)	0.00 (0.00)	0.01 (0.07)	0.01 (0.08)	0.324
Education: primary or less (share)	0.38 (0.49)	0.39 (0.49)	0.38 (0.49)	0.996
Education: secondary (share)	0.40 (0.49)	0.40 (0.49)	0.41 (0.49)	0.932
Education: tertiary+ (share)	0.17 (0.38)	0.19 (0.39)	0.18 (0.38)	0.804
Education: other/unknown (share)	0.04 (0.20)	0.03 (0.17)	0.03 (0.18)	0.648
Religion: Muslim (share)	0.90 (0.30)	0.92 (0.28)	0.91 (0.28)	0.657
Religion: non-Muslim (share)	0.09 (0.29)	0.08 (0.27)	0.08 (0.28)	0.788
Religion: other/unknown (share)	0.01 (0.09)	0.00 (0.05)	0.00 (0.07)	0.595
Ethnicity: Javanese (share)	0.02 (0.13)	0.02 (0.15)	0.02 (0.14)	0.935
Ethnicity: other (share)	0.98 (0.13)	0.98 (0.15)	0.98 (0.14)	0.935

Table A6: Sample characteristics compared to Indonesian population and social media users.

Categories	Sample	SM Users	Population
<u>Gender</u>			
Male	49.67%	49.80%	49.80%
Female	49.93%	50.20%	50.20%
<u>Age</u>			
18–24	12.98%	12.71%	17.90%
25–34	18.04%	24.29%	26.30%
35–44	24.57%	32.11%	22.40%
45–54	25.90%	19.48%	16.40%
55+	18.51%	9.94%	17.00%
<u>Region</u>			
Sumatera	22.84%	23.06%	20.40%
Java and Bali	56.52%	54.12%	61.10%
Central and Eastern Provinces	20.64%	22.82%	18.50%
<u>Religion</u>			
Islam	90.68%	89.89%	87.40%
Christian	6.39%	7.50%	9.30%
Others	2.93%	2.61%	2.80%
<u>Education</u>			
Less than High School	34.89%	42.22%	67.40%
High School	47.67%	41.08%	25.30%
Higher than High School	17.44%	16.71%	7.20%

Table A7: Descriptive statistics: follow-up survey (wave 3) respondents.

Measure	N	Mean	St. Dev.	Min.	Median	Max.
<i>Demographics</i>						
Age (years)	974	40.31	12.71	18	42	92
Gender: male (share)	974	0.47	0.50	0	0	1
Gender: female (share)	974	0.52	0.50	0	1	1
Gender: other/unknown (share)	974	0.00	0.06	0	0	1
Education: primary or less (share)	974	0.37	0.48	0	0	1
Education: secondary (share)	974	0.41	0.49	0	0	1
Education: tertiary+ (share)	974	0.18	0.39	0	0	1
Education: other/unknown (share)	974	0.03	0.17	0	0	1
Religion: Muslim (share)	974	0.91	0.28	0	1	1
Religion: non-Muslim (share)	974	0.09	0.28	0	0	1
Religion: other/unknown (share)	974	0.00	0.06	0	0	1
Ethnicity: Javanese (share)	974	0.02	0.15	0	0	1
Ethnicity: other (share)	974	0.98	0.15	0	1	1
<i>TV and media (FU)</i>						
TV: hours watched (FU)	940	2.78	0.91	1	3	5
TV: news consumption (FU)	931	2.82	0.96	1	3	5
TV: political news (FU)	939	2.35	1.05	1	2	5
Number of TV channels watched (FU)	974	3.09	2.75	0	2	22
<i>Religious organizations (FU)</i>						
Religious org: NU membership (FU)	940	1.66	1.04	1	1	4
Religious org: Muhammadiyah (FU)	940	1.42	0.86	1	1	4
Religious org: other Islamic (FU)	928	1.41	0.86	1	1	4

B.2 Instrumental Variables Design

Table A8: Descriptive statistics: instrumental variables sample.

Measure	N	Mean	St. Dev.	Min.	Median	Max.
<i>Governance</i>						
Internet access	141974	1.46	0.96	1	1	4
Presidential satisfaction	160231	0.41	1.11	-2	1	2
Corruption evaluation	63113	0.20	1.05	-2	1	2
Democracy satisfaction	109880	0.40	1.05	-2	1	2
<i>Intolerance</i>						
Intol: religious events	21675	0.60	0.49	0	1	1
Intol: worship building	30062	0.48	0.49	0	0	1
Intol: non-Muslim bupati	15457	0.41	0.49	0	0	1
<i>Religious participation</i>						
Religious org membership	124915	2.64	0.69	1	3	3
Pengajian frequency	102848	2.58	0.97	1	3	4
<i>Instruments & coverage</i>						
Elevation (std. dev.)	108109	40.17	79.10	0	8.89	900.53
Dist. nearest city (log)	108419	4.49	1.35	0.26	4.51	7.79
<i>Village-level controls (PODES 2014)</i>						
HH electricity share	108419	0.97	0.12	0	1	1
Travel time to bupati (hrs)	108419	1.56	3.07	1	1	98

C Additional Analyses

C.1 Experimental Design

C.1.1 Local Average Treatment Effect (LATE) Estimation

The following figures report local average treatment effect (LATE) estimates obtained via two-stage least squares, instrumenting actual compliance (averaging fewer than 5 minutes of daily social media use on each platform, based on device monitoring data) with treatment assignment. These complement the intent-to-treat estimates reported in the main text.

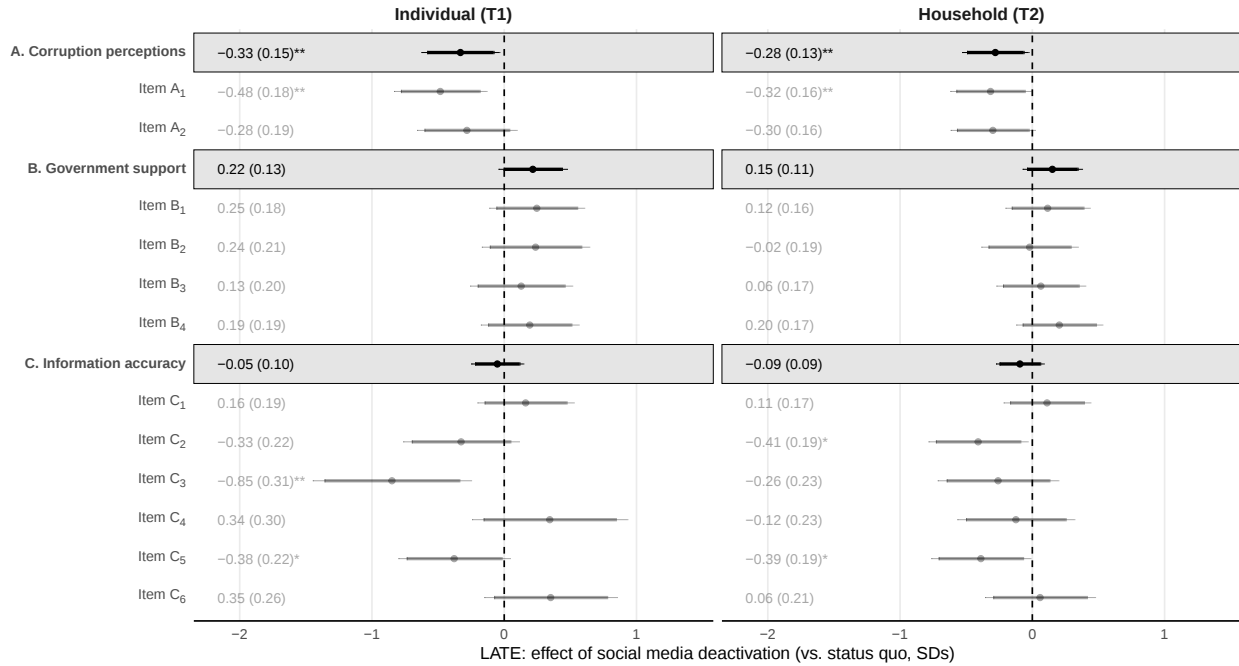


Figure A3—LATE estimates for government outcomes. Two-stage least squares estimates instrumenting compliance with treatment assignment. Thick bars denote 90% confidence intervals; thin bars denote 95% confidence intervals. Shaded rows indicate index-level estimates.

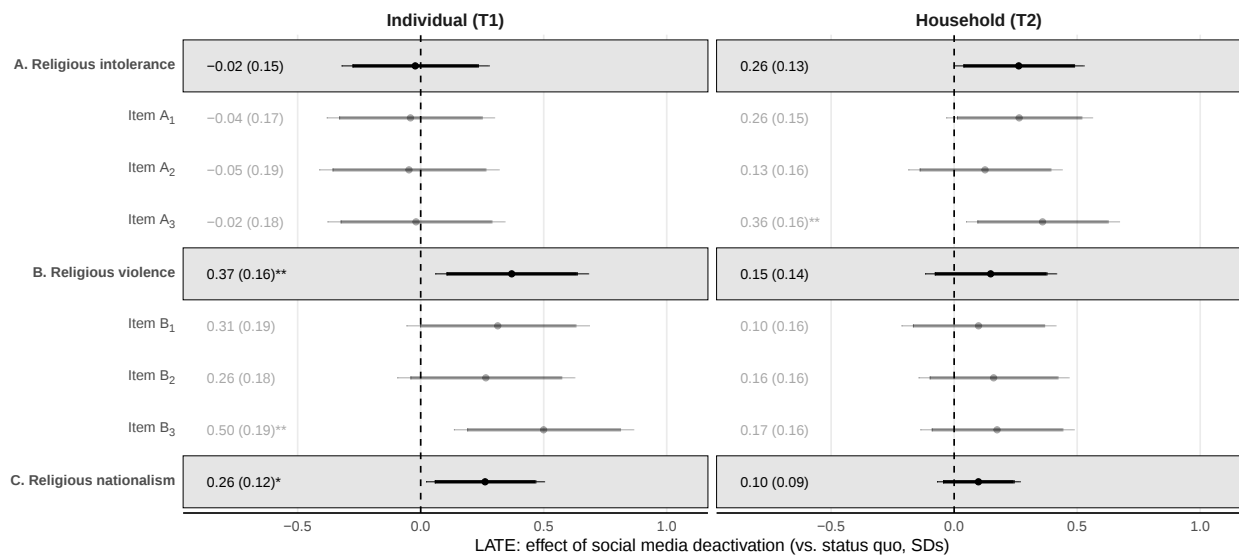


Figure A4—LATE estimates for religious outcomes (Muslim subsample). Two-stage least squares estimates instrumenting compliance with treatment assignment.

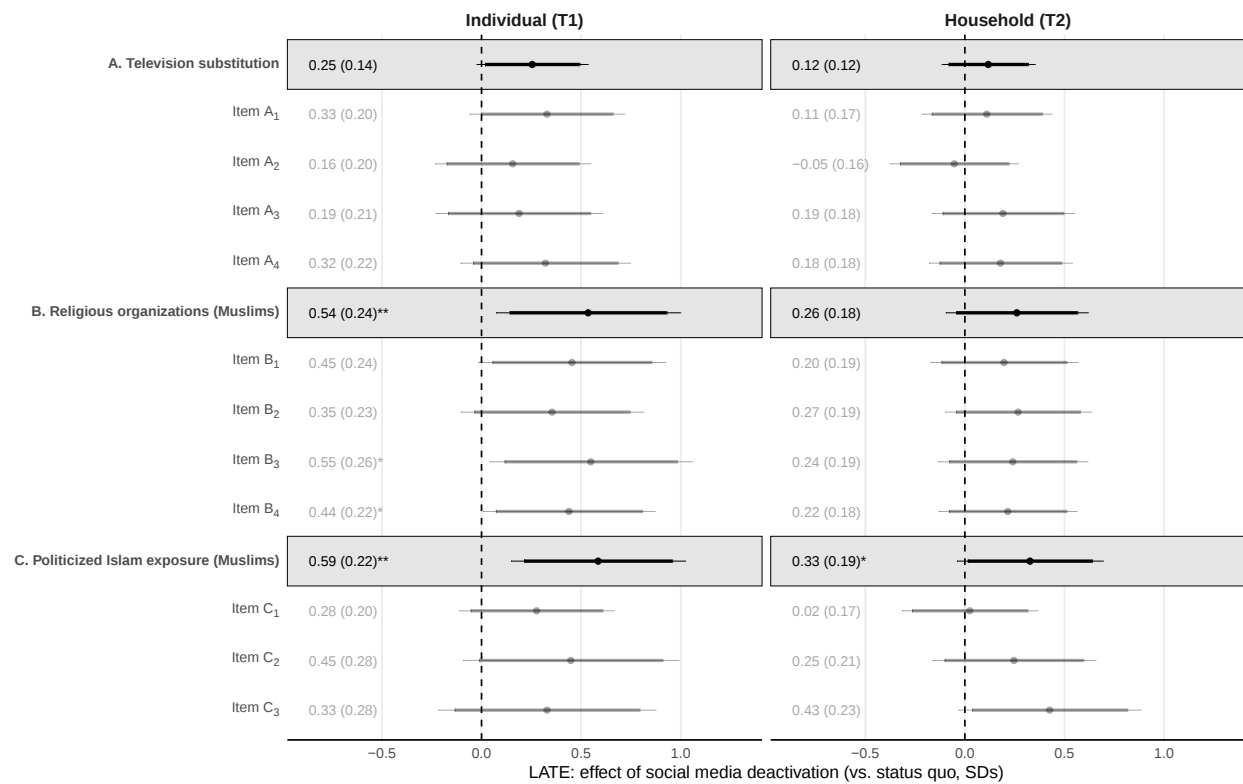


Figure A5—LATE estimates for follow-up survey outcomes. Two-stage least squares estimates instrumenting compliance with treatment assignment.

C.1.2 Specification with Pre-Treatment Controls

The main intent-to-treat estimates reported in the main text regress endline outcomes on treatment indicators and demographic controls, without conditioning on the pre-treatment (baseline) value of each outcome. As a robustness check, I re-estimate all specifications as ANCOVA models that additionally control for the baseline value of the dependent variable, $y_{i,0}$, following the recommendation of McKenzie (2012). Including the lagged dependent variable improves precision by absorbing individual-level variation in pre-treatment outcome levels.

Figures A6 and A7 present the results. The pattern of findings is substantively unchanged. For government outcomes, deactivation continues to reduce perceived corruption in both treatment arms ($\hat{\beta} = -0.13$ and -0.12 SDs, respectively), while government support and information accuracy remain null at the index level. For religious outcomes, the ANCOVA specification sharpens the estimates for religious violence in the individual arm ($\hat{\beta} = 0.14$, $SE = 0.07$) and religious nationalism ($\hat{\beta} = 0.12$, $SE = 0.05$), consistent with deactivation increasing support for religiously motivated violence and exclusionary conceptions of national identity among treated respondents.

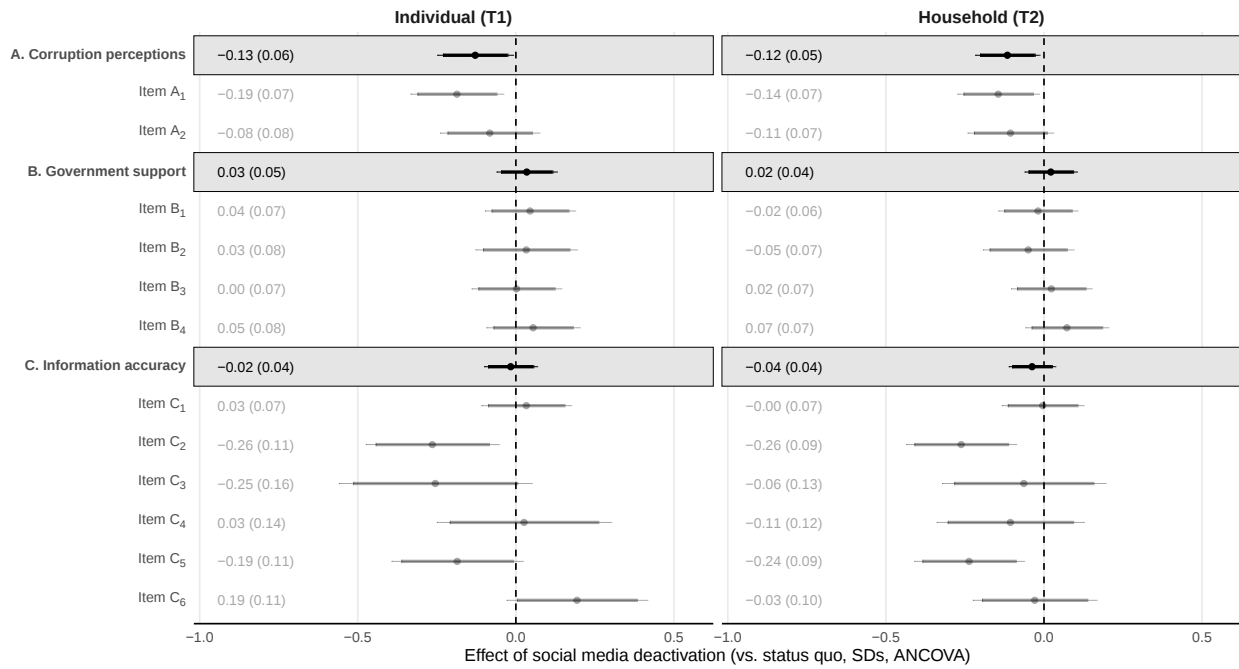


Figure A6—Government outcomes: specification with pre-treatment controls

Note: Intent-to-treat estimates from OLS regressions of endline outcomes on treatment indicators, controlling for the baseline value of the dependent variable and demographic covariates (gender, age, province, incentive level, religion, education, district). Thick bars denote 90% confidence intervals; thin bars denote 95% confidence intervals. Heteroskedasticity-robust standard errors. Shaded rows indicate index-level estimates; sub-items shown at reduced opacity. All outcomes are standardized using baseline control-group moments.

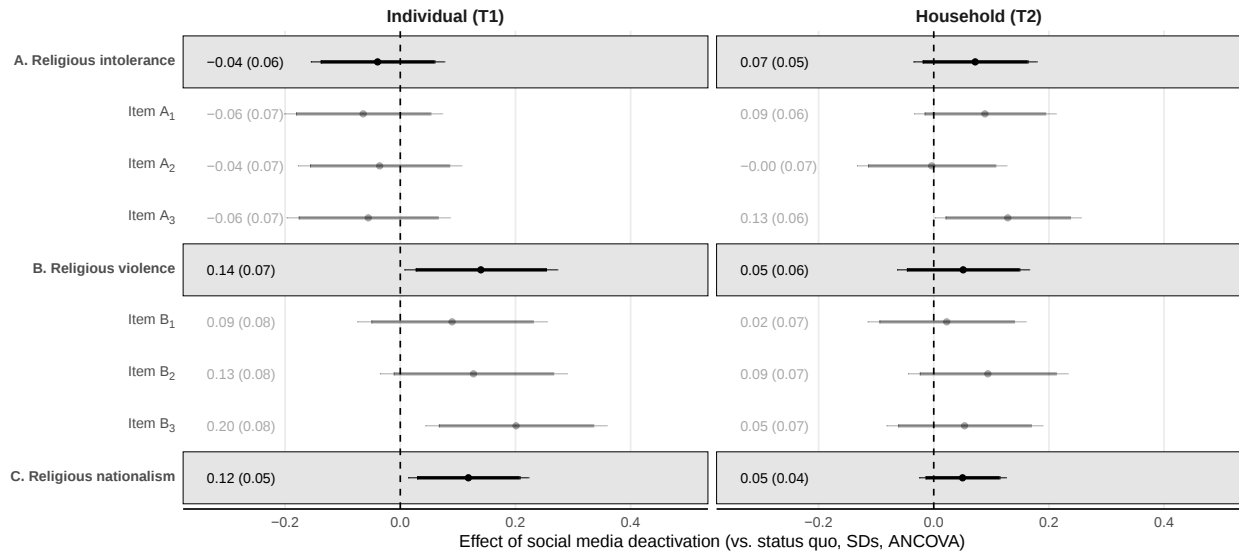


Figure A7—Religious outcomes: specification with pre-treatment controls (Muslim subsample)

Note: Intent-to-treat estimates from OLS regressions of endline outcomes on treatment indicators, controlling for the baseline value of the dependent variable and demographic covariates (gender, age, province, incentive level, education, district). Sample restricted to Muslim respondents. Thick bars denote 90% confidence intervals; thin bars denote 95% confidence intervals. Heteroskedasticity-robust standard errors. Shaded rows indicate index-level estimates; sub-items shown at reduced opacity. All outcomes are standardized using baseline control-group moments.

C.1.3 Raw Outcome Means by Treatment Group

Figure A8 presents unadjusted group means for each component item of the religious intolerance, religious violence, and religious nationalism outcomes at both baseline and endline. This figure complements the regression-based estimates in the main text by displaying the raw data without any standardization or covariate adjustment. At baseline, the three treatment arms are well-balanced across all items, as expected from randomization. At endline, the treatment arms—particularly the individual deactivation arm—show higher mean levels of support for religious violence and religious nationalism relative to the control group. The intolerance items show a more mixed pattern, with the household arm exhibiting modest increases while the individual arm remains closer to the control.

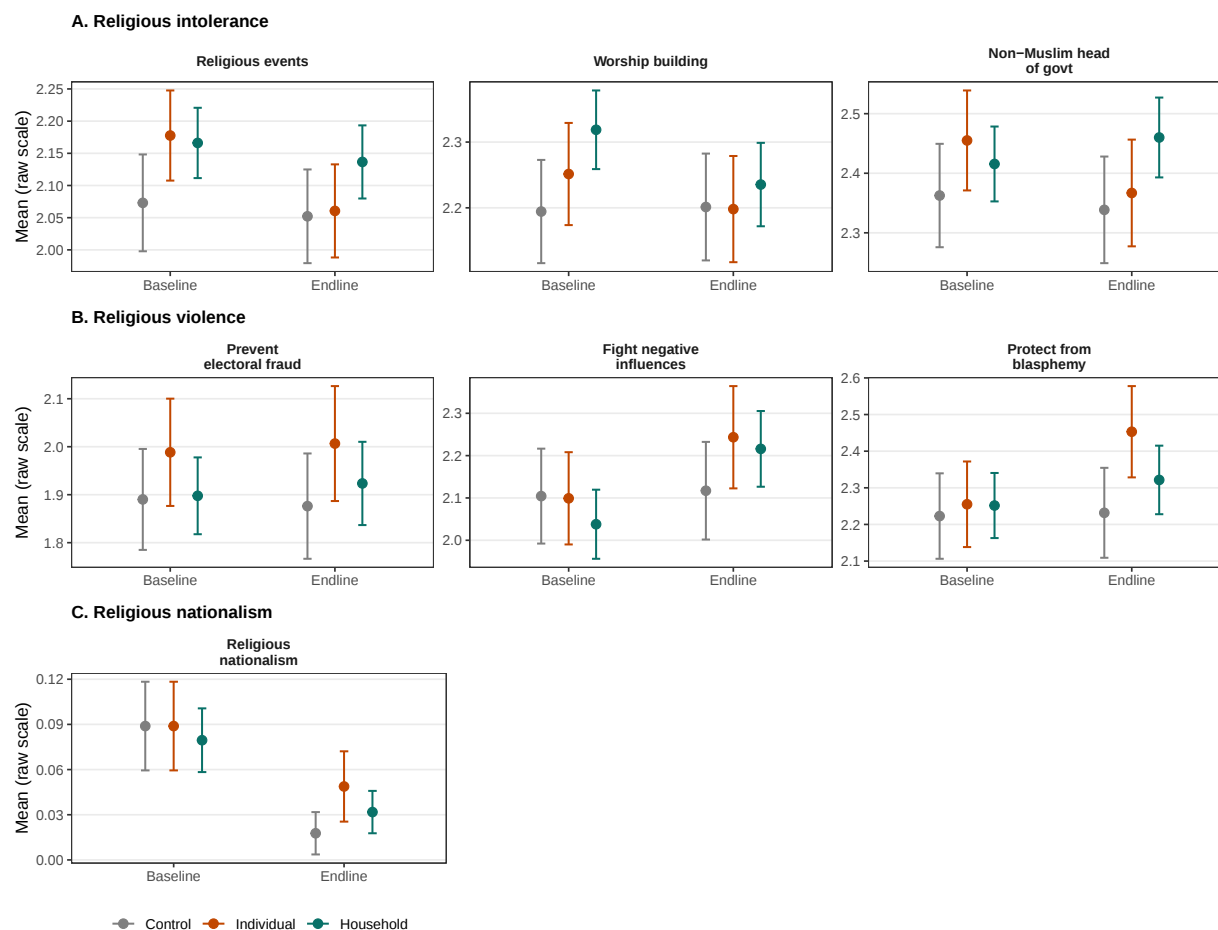


Figure A8—Raw means of religious outcome items by treatment group and wave

Note: Unadjusted group means with 95% confidence intervals, computed separately by treatment arm and survey wave. Sample restricted to Muslim respondents. Panel A: Non-Muslim intolerance items (1–4 scale; higher values indicate greater intolerance). Panel B: Support for religious violence items (1–4 scale; higher values indicate greater justification of violence). Panel C: Religious nationalism item (binary; 1 = respondent identifies being Muslim as the most important aspect of being Indonesian). All items have been cleaned (non-response codes recoded to NA) and, where applicable, reversed so that higher values indicate more of the named construct.

Figure A9 presents the analogous exercise for governance outcomes. At baseline, the three treatment arms are well-balanced across all items. At endline, the corruption perception items show modest declines in the treatment arms relative to the control group, consistent with the regression-based estimates. Government support items and information accuracy items remain largely unchanged across arms, with no visually discernible separation between treatment and control.

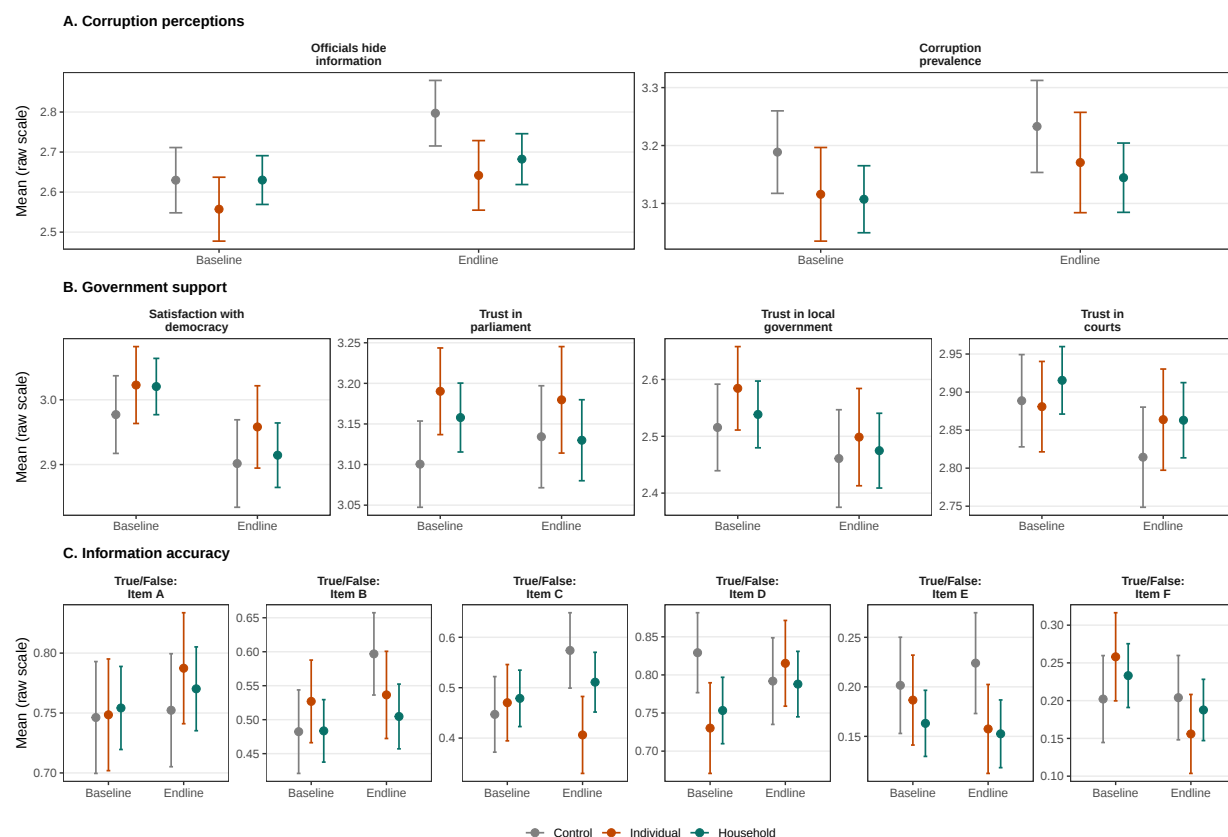


Figure A9—Raw means of governance outcome items by treatment group and wave

Note: Unadjusted group means with 95% confidence intervals, computed separately by treatment arm and survey wave. Full sample. Panel A: Corruption perception items (1–4 scale; higher values indicate greater perceived corruption). Panel B: Government support items (1–4 scale; reversed so that higher values indicate greater satisfaction or trust). Panel C: Information accuracy items (binary; 1 = correct answer to true/false political knowledge question). All items have been cleaned (non-response codes recoded to NA) and, where applicable, reversed so that higher values indicate more of the named construct.

C.1.4 Tabular Regression Estimates

Tables A9, A10, and A11 present intent-to-treat regression estimates corresponding to the coefficient plots in Figures 3, 4, and 5, respectively. All outcomes are standardized using baseline control-group moments.

Table A9: ITT estimates: governance outcomes.

	Dependent variable		
	Corruption	Gov. support	Info. accuracy
	(1)	(2)	(3)
Individual deactivation (T1)	-0.116* (0.060)	0.033 (0.048)	-0.017 (0.042)
Household deactivation (T2)	-0.114** (0.052)	0.020 (0.042)	-0.040 (0.037)
Baseline outcome	0.382*** (0.030)	0.599*** (0.030)	0.263*** (0.034)
Female	0.051 (0.044)	-0.039 (0.036)	-0.011 (0.031)
Age	-0.004** (0.002)	0.003** (0.001)	0.002 (0.001)
Incentive = 2	-0.013 (0.054)	0.068 (0.043)	-0.014 (0.039)
Incentive = 3	-0.023 (0.054)	0.091** (0.044)	-0.011 (0.039)
<i>N</i>	1,327	1,397	1,131
<i>R</i> ²	0.192	0.323	0.138
Religion FE	✓	✓	✓
Education FE	✓	✓	✓
District FE	✓	✓	✓

Notes: Intent-to-treat estimates. Heteroskedasticity-robust standard errors in parentheses. All outcomes standardized using baseline control-group moments. Reference category: Control group. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A10: ITT estimates: religious outcomes (Muslim subsample).

	Dependent variable		
	Intolerance	Violence	Nationalism
	(1)	(2)	(3)
Individual deactivation (T1)	-0.019 (0.058)	0.132** (0.067)	0.117** (0.052)
Household deactivation (T2)	0.061 (0.054)	0.056 (0.058)	0.055 (0.039)
Baseline outcome	0.400*** (0.034)	0.299*** (0.031)	0.060** (0.029)
Female	-0.028 (0.043)	0.062 (0.049)	-0.011 (0.039)
Age	0.002 (0.002)	-0.003 (0.002)	0.000 (0.001)
Incentive = 2	-0.001 (0.054)	-0.024 (0.061)	0.019 (0.044)
Incentive = 3	0.003 (0.053)	-0.032 (0.063)	0.027 (0.045)
<i>N</i>	1,203	1,180	1,254
<i>R</i> ²	0.240	0.149	0.039
Education FE	✓	✓	✓
District FE	✓	✓	✓

Notes: Intent-to-treat estimates. Sample restricted to Muslim respondents. Heteroskedasticity-robust standard errors in parentheses. All outcomes standardized using baseline control-group moments. Reference category: Control group. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A11: ITT estimates: follow-up survey outcomes.

	Dependent variable		
	TV subst.	Relig. orgs	Polit. Islam
	(1)	(2)	(3)
Individual deactivation (T1)	0.118* (0.064)	0.264** (0.103)	0.263*** (0.100)
Household deactivation (T2)	0.056 (0.054)	0.112 (0.085)	0.142 (0.087)
Female	0.010 (0.046)	-0.086 (0.075)	-0.301*** (0.076)
Age	0.005** (0.002)	-0.009*** (0.003)	0.002 (0.003)
Incentive = 2	0.054 (0.057)	-0.017 (0.093)	0.009 (0.092)
Incentive = 3	0.049 (0.056)	0.012 (0.094)	-0.125 (0.092)
<i>N</i>	974	872	824
<i>R</i> ²	0.068	0.091	0.089
Religion FE	✓	✓	✓
Education FE	✓	✓	✓
District FE	✓	✓	✓

Notes: Intent-to-treat estimates. Column 1 uses the full sample; Columns 2–3 restricted to Muslim respondents. No baseline control is available for follow-up outcomes. Heteroskedasticity-robust standard errors in parentheses. Reference category: Control group. Religion FE included only in Column 1 (full sample). * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

C.1.5 Alternative Explanations

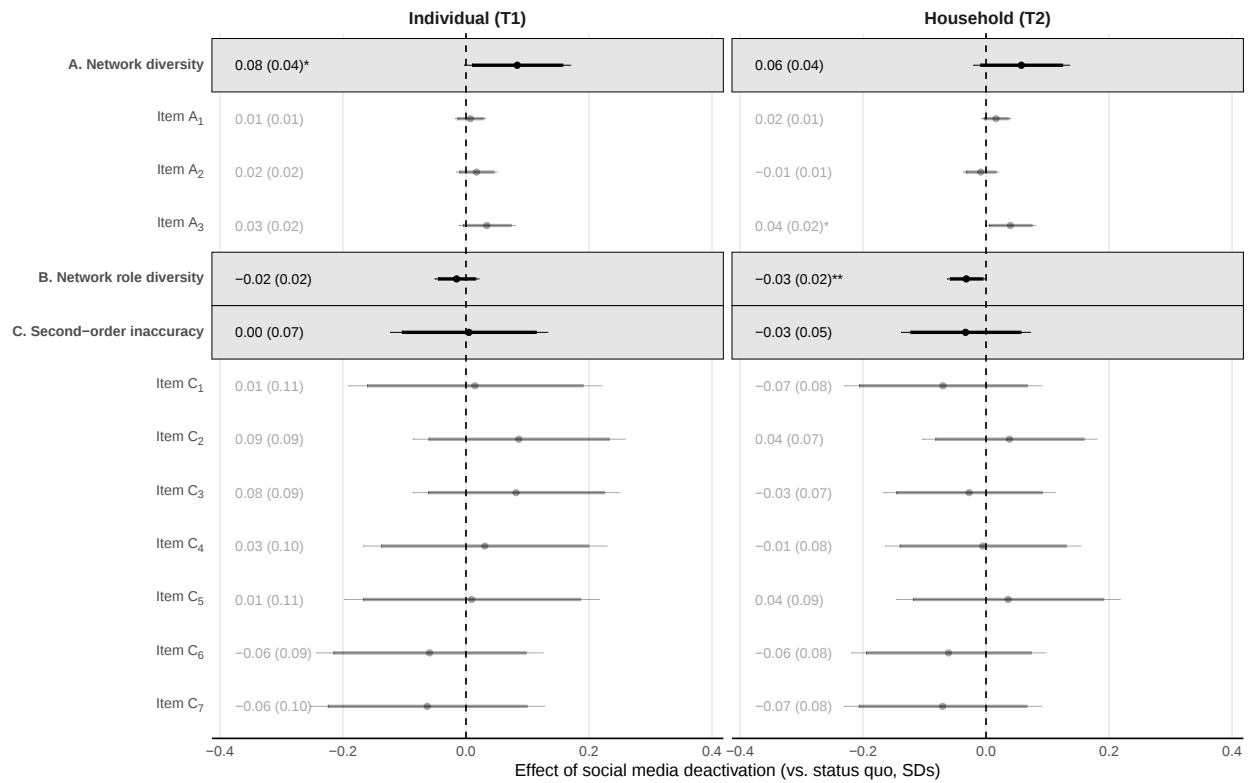


Figure A10—Effects of social media deactivation on network diversity and second-order beliefs. Point estimates are ITT coefficients with 90% (thick) and 95% (thin) confidence intervals. Item A₁: Religious diversity of social network (time-weighted). Item A₂: Ideological diversity of social network (time-weighted). Item A₃: Ethnic diversity of social network (time-weighted). Panel B: Role diversity of social network (share of contacts across distinct relationship types). Panel C: Second-order inaccuracy index; respondents estimated the percentage of various social groups in Indonesia, and each item is scored as squared deviation from the external benchmark. Items C₁–C₇: Ahmadiyah adherents (benchmark: 0.1%), Prabowo supporters (58.6%), Javanese (40%), Chinese-Indonesians (1.2%), FPI supporters (33%), violent crime victims (1%), domestic violence victims (25%).

C.2 Instrumental Variables Design

C.2.1 Robustness: Distance-to-City Controls

A potential concern with the terrain-based instruments is that rugged terrain and distance from major cities may correlate with remoteness from economic and administrative centers, which could independently affect political attitudes through channels unrelated to internet access. To address this, I re-estimate all IV specifications adding the log distance from the village centroid to Jakarta as a control variable, which proxies for proximity to major economic and administrative centers. Table A12 shows that the main results are robust to this additional control.

Table A12: IV Robustness: Adding Distance-to-City Controls

	Government outcomes			Religious tolerance outcomes		
	Pres. Satis.	Corrup. Eval.	Dem. Satis.	Intol. Events	Intol. Worship	Intol. Bupati
	(1)	(2)	(3)	(4)	(5)	(6)
Internet	-0.495*** (0.100)	-0.328** (0.149)	-0.416*** (0.082)	-0.284*** (0.057)	-0.365*** (0.096)	-0.154*** (0.052)
Num.Obs.	67517	21418	42114	15019	12778	13101
R2	-0.036	0.053	-0.033	-0.136	-0.284	0.001
Year FE	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓
City dist. control	✓	✓	✓	✓	✓	✓

Notes: This table re-estimates the main IV specifications from the text, adding the log distance from the village centroid to Jakarta as a control for proximity to major economic and administrative centers. Odd-numbered columns reproduce the baseline specification; even-numbered columns add the distance control. The outcome in Columns 1–6 is indicated by the column header; all intolerance outcomes are reverse-coded so that higher values indicate greater intolerance. The endogenous variable is self-reported internet access (1–4 scale), instrumented by terrain ruggedness (standard deviation of elevation, in km) and log distance to the nearest major city. All specifications include year and province fixed effects. Standard errors clustered at the district level are reported in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

C.2.2 Robustness: Village-Level Economic Development Controls

A related concern is that terrain characteristics may be correlated with village-level economic development, which could affect political attitudes independently of internet access. For example, villages with flatter terrain may be both easier to equip with cell towers and more economically developed, generating a spurious association between mobile internet coverage and political outcomes. To address this, I supplement the IV specifications with two village-level controls drawn from the 2014 Village Census (PODES): (1) the share of households with electricity (PLN or non-PLN), which proxies for the overall level of economic infrastructure and modernization; and (2) the travel time in hours from the village to the district capital (*kantor bupati/walikota*), which captures administrative remoteness and access to government services. I add these controls individually and jointly to assess whether the estimated effect of internet access is sensitive to conditioning on village-level economic conditions.

Tables A13 and A14 present the results. Across all outcomes, the estimated coefficients on instrumented internet access are virtually unchanged by the inclusion of the PODES controls, whether entered individually or together. This stability suggests that the terrain-based instruments are not simply proxying for village-level economic development.

Table A13: IV Robustness: Religion and Tolerance Outcomes with PODES Economic Controls

	Intolerance (reversed)									Religious practice					
	Relig. Events			Worship Bldg.			Non-Muslim Bupati			Relig. Org			Relig. Study		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Internet	-0.291*** (0.053)	-0.290*** (0.051)	-0.291*** (0.053)	-0.364*** (0.088)	-0.357*** (0.085)	-0.364*** (0.089)	-0.131*** (0.045)	-0.131*** (0.044)	-0.132*** (0.046)	-0.129*** (0.036)	-0.119*** (0.034)	-0.130*** (0.036)	-0.229* (0.120)	-0.203* (0.116)	-0.234* (0.121)
Observations	15,019	15,019	15,019	12,778	12,778	12,778	13,095	13,095	13,095	49,401	49,401	49,401	43,082	43,082	43,082
R-squared	-0.152	-0.150	-0.152	-0.281	-0.265	-0.282	0.025	0.025	0.024	-0.040	-0.023	-0.041	0.010	0.015	0.009
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Electricity share	✓		✓	✓		✓	✓		✓	✓		✓	✓		✓
Travel to bupati		✓	✓		✓	✓		✓	✓		✓	✓		✓	✓

Notes: This table re-estimates the IV specifications for religion and tolerance outcomes, progressively adding village-level economic controls from the 2014 Village Census (PODES). For each outcome, Column 1 reproduces the baseline specification. Column 2 adds the village-level share of households with electricity. Column 3 adds the travel time (hours) from the village to the district capital. Column 4 includes both controls simultaneously. Intolerance outcomes are reverse-coded so that higher values indicate greater intolerance. The endogenous variable is self-reported internet access (1–4 scale), instrumented by terrain ruggedness (standard deviation of elevation, in km) and log distance to the nearest major city. All specifications include year and province fixed effects. Standard errors clustered at the district level are reported in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A14: IV Robustness: Government Outcomes with PODES Economic Controls

	Second stage: Government outcomes (IV)								
	Pres. Satisfaction			Corruption Eval.			Democracy Satis.		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Internet	-0.451*** (0.089)	-0.448*** (0.086)	-0.447*** (0.090)	-0.273** (0.133)	-0.268** (0.131)	-0.264* (0.135)	-0.407*** (0.074)	-0.408*** (0.072)	-0.406*** (0.075)
Observations	67,517	67,517	67,517	21,418	21,418	21,418	42,114	42,114	42,114
R-squared	-0.017	-0.016	-0.015	0.068	0.069	0.070	-0.028	-0.029	-0.028
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Electricity share	✓		✓	✓		✓	✓		✓
Travel to bupati		✓	✓		✓	✓		✓	✓

Notes: This table re-estimates the IV specifications for government outcomes, progressively adding village-level economic controls from the 2014 Village Census (PODES). For each outcome, Column 1 reproduces the baseline specification. Column 2 adds the village-level share of households with electricity. Column 3 adds the travel time (hours) from the village to the district capital. Column 4 includes both controls simultaneously. The endogenous variable is self-reported internet access (1–4 scale), instrumented by terrain ruggedness (standard deviation of elevation, in km) and log distance to the nearest major city. All specifications include year and province fixed effects. Standard errors clustered at the district level are reported in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.