Reducing Social Media Usage During Elections: Evidence from a Multi-Country WhatsApp Experiment

Rajeshwari Majumdar[*]
Tiago Ventura[†]
Shelley Liu[‡]
Carolina Torreblanca[§]
Joshua A. Tucker[¶]

December 11, 2025

Abstract

Social media messaging platforms are central to worldwide communication, but are also major hubs of misinformation and toxic content. On these platforms, information spreads through interpersonal and group-based chats rather than feed-based recommendations. We argue that introducing barriers to usage can increase the costs of consuming low-quality content and promote more deliberate engagement, shaping information consumption and downstream attitudes. We evaluate our argument through three coordinated online field experiments in Brazil, India, and South Africa. We incentivize participants to either avoid multimedia content on WhatsApp or to limit their usage to 10 minutes per day for four weeks ahead of each country's elections. Our interventions significantly reduced participants' exposure to uncivil political discussions and misinformation—but at the expense of keeping up with political news. However, political attitudes did not shift, although treated participants did report improved wellbeing, particularly when they substituted WhatsApp usage with more offline activities.

Keywords. Social Media, Global South, WhatsApp, Misinformation, Polarization, South Africa, India, Brazil

[*] School of Media and Public Affairs, The George Washington University. r.majumdar@gwu.edu.
[†] McCourt School of Public Policy, Georgetown University. tv186@georgetown.edu.
[‡] Sanford School of Public Policy, Duke University. shelley.liu@duke.edu.
[§] PDRI-DevLab, University of Pennsylvania. catba@sas.upenn.edu.
[¶] Department of Politics and Center for Social Media and Politics, New York University. joshua.tucker@nyu.edu.

1 Introduction

Social media has reshaped political communication, influencing not only how information is disseminated but also how individuals engage with politics. Early on, these platforms sparked genuine enthusiasm among policymakers and experts about breaking informational barriers, facilitating connections, and mobilizing citizens (Tucker et al. 2017). However, these platforms have been increasingly associated with possible negative externalities. This issue is particularly salient in politics, where social media usage has been widely conceived as having deleterious consequences for democracy (Lorenz-Spreen et al. 2022; Budak et al. 2024).

Scholarship examining the political effects of social media, however, faces several limitations. First, much of our existing knowledge come from studies of traditional, feed-based social media platforms such as Facebook or X, which are particularly popular in the United States and other Western democracies. Yet in much of the world, information exchange and daily life are intertwined largely through social media messaging apps such as WeChat, Telegram, and—with over 2 billion active users globally—WhatsApp. It is not clear whether lessons from existing studies transfer well to these messaging apps because, unlike algorithmic feeds, content propagation on social messaging apps depends more heavily on users forwarding content in group chats and one-to-one conversations (Valenzuela, Bachmann and Bargsted 2021; Rossini et al. 2021). The most viral posts are therefore crafted for easy sharing across groups and chats, making its information environment dominated not by text-based news and creator posts, but by easily shareable, quasi-anonymous multimedia content such as videos, images, and audio (Ventura et al. 2025; Resende et al. 2019; Garimella and Eckles 2020).

Second, developing contexts often face different challenges associated with misinformation and information flow, which may generate different consumption patterns and thus downstream political and non-political outcomes. These may be capacity-driven, such as less robust fact-checking apparatuses or greater mobile data costs, which decreases the likelihood of consulting multiple information sources outside of social media (Bowles, Larreguy and Liu 2020; Haque et al. 2020). They may also be political in nature, considering increased risks of violence, low political trust, and greater political control over media in these countries (Jones 2022; Badrinathan, Chauchard and Siddiqui 2024). In such settings, the potential benefits of tackling misinformation on social media may therefore be larger. In response, scholars have increasingly sought to diversify research contexts (e.g., see Blair et al. 2024). However, these efforts often focus more narrowly on interventions to stymie the spread of misinformation and the persistence of inaccurate beliefs rather than on understanding the impacts of popular social media platforms on information flows more broadly (Badrinathan and Chauchard 2023; Bowles et al. 2025).

Taken together, these observations highlight the need for causal evidence on the political effects of social media in the Global South. Yet such research would be difficult to carry out using existing research designs developed for Western, feed-based platforms. Existing experimental studies have relied predominantly on deactivation designs (Allcott et al. 2020, 2024; Arceneaux et al. 2023; Asimovic et al. 2021; Asimovic, Nagler and Tucker 2023), which incentivize users to completely exit from feed-based platforms like Facebook for a set duration. Given messaging platforms' deep embeddedness in individuals' lives and daily tasks (Newman et al. 2024), a full deactivation is both impractical for research and impossible to implement in real-world policy. Thus, the dominant research design used to identify the causal effects of social media usage offers limited leverage in the Global South.[1]

To address these gaps, we investigate whether *reducing* WhatsApp usage during elections, when misinformation circulates widely and might shape political outcomes in the short run, can minimize the negative effects of social media in Global South contexts. We theorize that, in the context of social messaging apps such as WhatsApp, introducing small barriers to usage can foster more deliberate and selective engagement, shaping both political and social outcomes. Increased barriers to usage may encourage users to remain connected while being more intentional about which groups they visit and which messages they open. By reorienting users away from large, impersonal, and low-quality spaces, barriers to usage can reshape users' online information environment. In turn, this should reduce exposure to harmful political content and misinformation circulating on WhatsApp (especially during election periods), with downstream effects on political attitudes and overall well-being.

[1] A full deactivation may restrict participation to a non-representative subset of users, limiting what can be learned. It also raises ethical concerns as fully removing people from WhatsApp, where they communicate with family and conduct business, could impose undue costs.

We further note that this strategy and subsequent expectations are more applicable to social messaging apps like WhatsApp. In feed-based platforms like Facebook and X, reducing usage does not guarantee that users will be more selective and change their informational environment, as users' information exposure is predominantly controlled by algorithmic recommender systems. In contrast, information diffusion in messaging apps occurs primarily through interpersonal and group-based chats, giving users more control over what they consume.

To evaluate our argument empirically, we implement a large-scale online field experiment to reduce WhatsApp usage in three major Global South democracies—India, South Africa, and Brazil—for four weeks before elections in each of these countries. As three of the largest democracies in Asia, Africa, and Latin America, respectively, they are substantively important cases to study. Across these countries, and in most of the Global South, WhatsApp plays a major role in everyday life as the largest social media and messaging platform: 90% of the Brazilian, 71% of the South African, and 50% of the Indian adult population use WhatsApp daily (Poushter 2024) ahead of other apps such as Facebook, X, and YouTube. In all three countries, WhatsApp groups are heavily used to mobilize citizens around social, political, and identity issues (Chauchard and Garimella 2022; Gil de Zúñiga, Ardèvol-Abreu and Casero-Ripollés 2021; Kalogeropoulos and Rossini 2025). WhatsApp is also a breeding ground for misinformation. For example, in Brazil, during the height of the COVID-19 pandemic, voters listed WhatsApp as the main source of misinformation they saw about vaccines (Newman et al. 2021). In India, content demonizing minority ethnic groups proliferates on the platform, particularly during election periods (Chauchard and Garimella 2022; Saha et al. 2021). In South Africa, WhatsApp has become a pivotal channel for the circulation of rumors about politicians and vote-fraud claims (Bowles et al. 2025; Allen 2021). In sum, WhatsApp is not only the most used app for diverse tasks related to personal communications and business, but also a central hub for political information, where news, misinformation, and polarizing content routinely reach citizens.

Our design incentivizes a reduction in WhatsApp usage through two distinct interventions, each structured to test the mechanisms outlined in our theory. First, as misinformation and polarizing content on WhatsApp is often delivered through multimedia content, we incentivize one group of treated participants to turn off the automatic download of multimedia content (such as images and videos) on WhatsApp and not access any multimedia on WhatsApp during the study period (*Multimedia* arm). Second, we incentivize another group to limit their WhatsApp usage to 10 minutes per day during the same four weeks (*Time* arm). Taken together, our experiment aims to reduce participants' WhatsApp usage during election season and promote more deliberate engagement with content on the platform by introducing barriers, both through direct exposure to multimedia content and indirect exposure through time spent in the platform.

We present four core findings. First, as hypothesized, reducing WhatsApp usage decreased participants' reported exposure to online incivility and to toxic political discussions. Correspondingly, reducing WhatsApp usage consistently reduced participants' exposure to misinformation rumors—but at the expense of keeping up with true news circulating in the weeks before elections. Second, contrary to our pre-registered theoretical expectations, these shifts in information consumption neither increased skepticism nor improved discernment: our treatments did not influence participants' accuracy judgments for either misinformation or true news. Third, despite reducing treated participants' exposure to uncivil content and (mis)information about the elections, our treatments had no downstream effects on political attitudes, including affective (partisan) polarization, identity-based prejudice, issue polarization, and candidate favorability. At the same time, we uncover positive non-political effects, including a significant boost in subjective well-being. In sum, our intervention did not enhance political knowledge or reduce polarization, but in line with our theory of more deliberate engagement leading to changes in information consumption, it did yield considerable benefits in terms of decreasing exposure to low-quality content (including misinformation) and improving well-being.

Within the broader conversation about social media usage and its downstream effects in the Global South (Budak et al. 2024; Badrinathan, Chauchard and Siddiqui 2024), our novel multi-country design also allows us to explore how both country context and individual consumption patterns shape responses to social media reduction. We document two notable findings. First, we observe reductions in news recall primarily in Brazil and South Africa, and not in India; second, we observe larger and more precisely estimated reductions in online toxicity and improvements in subjective well-being in Brazil. This may be because Indian participants reported relatively less reliance on WhatsApp for news at baseline, explaining the null effects on news recall, while Brazilian participants were relatively more likely to substitute toward offline activities rather than other social media platforms, potentially amplifying gains in well-being. Together, these results suggest that reductions in social media usage have larger positive effects when they also prompt a shift to different offline behaviors. These country-level differences complement our heterogeneous effects analyses, where we find that the impact of reduction appears strongest among heavy WhatsApp users and those with weaker partisan identities—indicating that both baseline consumption and engagement patterns condition the effects of social media reduction.

Our study contributes both substantively and methodologically to existing scholarship on the effects of social media. Substantively, as our study was deployed one month prior to major elections in each country, it speaks to research on media, information access, and political attitudes during politically polarizing periods. Elevated polarization reduces individuals' willingness to update their priors, increase the salience of directional motivations and identity or partisan cues (Graham and Svolik 2020; Druckman, Peterson and Slothuus 2013). Considering the media environment, this dynamic leads voters to increase their overall demand for political content, as well as their reliance on pro-attitudinal sources (Arceneaux and Johnson 2013; Stroud 2011). In this context, while elections are a critical period in democratic politics, these are also times in which political attitudes become increasingly more resistant to change. Our null effects on a wide range of political attitudes, despite considerable reductions in exposure to (mis)information and uncivil political content, align closely with theories of media minimal effects in polarizing times and resonate with findings from other recent social media field experiments deployed during elections (Allcott et al. 2024; Nyhan et al. 2023; Guess et al. 2023; Ventura et al. 2025). Our heterogeneous effects across three settings, however, help to explain when and why reductions in social media use may or may not shift political outcomes.

Methodologically, our study expands the emerging literature on the causal effects of social media (Allcott et al. 2020, 2024; Arceneaux et al. 2023; Asimovic et al. 2021; Asimovic, Nagler and Tucker 2023). To date, most of these studies have completely deactivated Facebook users and maintained a strong focus on Western countries, with the exception of Ventura et al. (2025). Yet in many Global South democracies, messaging apps such as WhatsApp have become the central channel for misinformation and inflammatory content, while also remaining the central channel for personal and professional communications. Considering WhatsApp's deep integration into everyday life, our study instead pursues usage reduction. By examining novel information environments, differing consumption patterns, and downstream political and non-political attitudes across three major Global South democracies, our study broadens the scope of social media research and provides new comparative insights.

2 Reducing WhatsApp Usage

The potential impact of social media platforms on citizens' political attitudes and behaviors is often linked to how these platforms may facilitate the dissemination of low-quality, polarizing, or entirely false content (Guess, Nyhan and Reifler 2018; Flaxman, Goel and Rao 2016). Such concerns are particularly pronounced during periods of high political polarization, including election seasons. A rapidly expanding body of research has explored the broad societal implications of social media use, including its role in exacerbating political polarization (Banks et al. 2021; Settle 2018), distorting beliefs about the veracity of information (Pennycook, Cannon and Rand 2018; Anspach and Carlson 2020), deepening policy divisions (Velez and Liu 2024), and negatively affecting mental health, especially among younger users (Vanman, Baker and Tobin 2018; Hanley, Watt and Coventry 2019).

These concerns have motivated a growing experimental literature that identifies the causal effects of social media use through deactivation designs, wherein treated participants refrain from using social media (specifically, Facebook and/or Instagram) for a brief period (Allcott et al. 2020, 2024; Asimovic et al. 2021; Asimovic, Nagler and Tucker 2023; Arceneaux et al. 2023). These studies generally find reductions in exposure to (mis)information and increases in well-being, but no effects on downstream political outcomes such as political participation and polarization.[2] Yet, these approaches are difficult to scale, especially in contexts where social media platforms serve as indispensable communication tools. This limitation is particularly salient in the Global South, where messaging apps like WhatsApp function as both a social media platform and an essential communication tool for daily life.

We thus consider alternative interventions and theorize that simply *reducing*, rather than eliminating, WhatsApp use can meaningfully change how users consume information and thereby shape political and social outcomes. We argue that, by raising the costs of engagement, social me-

[2] Exceptions include Allcott et al. (2020), which finds that Facebook deactivation reduced affective polarization, and Allcott et al. (2024), which finds suggestive evidence of the same.

dia reduction fosters more deliberate engagement with content on the platform. In the context of messaging apps that are not centered around feed-based recommender systems, more deliberate engagement may change users' informational environment, producing, for example, reductions in exposure to false content, toxic discussions, and polarizing rhetoric, and have consequential effects on downstream political attitudes and well-being. This intuition is further developed as we present our pre-registered hypotheses in this section.

We employ two complementary interventions to reduce WhatsApp usage. First, we instruct some users to reduce usage by turning off their automatic *multimedia downloads* and not consume any multimedia content. WhatsApp automatically downloads multimedia by default, meaning that individuals' phones receive all images, videos, and documents (including forwarded ones). Turning off this function means all multimedia received is blurred, requiring users to manually choose and download the pieces of multimedia they would like to view. By introducing an additional decision point and asking users not to consume multimedia, this form of usage reduction narrows exposure to unsolicited content. Furthermore, as some users will unavoidably consume some multimedia content, simply turning off multimedia downloads in the first place incentivizes them to be strategic about which multimedia content to open. Our second approach to reducing usage is to require users to limit the *time* spent on the platform itself. This approach focuses on the overall intensity of engagement: by spending less time on WhatsApp, users also reduce the volume of interaction and content that they encounter. While the latter approach arguably imposes tighter behavioral constraints, both approaches meaningfully limit usage without forcing users to fully disconnect from WhatsApp, an essential tool for everyday communication.

2.1 How Reducing Usage Affects Online Information Consumption

Limiting usage on platforms like WhatsApp, while still allowing individuals to use these platforms for their daily communication needs, can meaningfully shape participants' informational environments. Specifically, usage reduction fosters more deliberate engagement with the platform and changes users' online information consumption through two primary channels: (1) prioritizing more personal and relevant communication spaces over larger, impersonal ones, and (2) exercising greater selectivity and attention toward the content they consume.

If participants are prompted to reduce WhatsApp usage, we can reasonably expect that they would likely focus their limited time towards connecting with close networks (such as friends and family) or focusing on work-related communications. Such prioritization should lead participants to reduce the amount of time they spend on consuming content from larger, anonymous groups or broadcast channels, where low-quality political discussions are more likely to occur. Along the same lines, decreasing WhatsApp usage can limit exposure to uncivil and intolerant speech (Rossini 2022). Large, impersonal groups and channels are frequent breeding grounds for toxic discourse: they tend to amplify extreme voices (Bor and Petersen 2022) while anonymity and social distance encourage negative behaviors that are less common in more personal settings (Rowe 2015). By curbing time in these spaces, users are also distancing themselves from the incivility that pervades broader online discussions.

H1: Reducing WhatsApp usage decreases exposure to low-quality political discussions.
H2: Reducing WhatsApp usage decreases exposure to uncivil political content.

Reducing WhatsApp usage should also reduce exposure to misinformation through several pathways. First, large group chats and channels are especially fertile ground for the circulation of viral, low-quality content, including false or misleading information (Anspach and Carlson 2020). Thus, by prioritizing against these spaces, individuals ought to be exposed to less misinformation. Second, even within personal networks, misinformation frequently spreads organically as viral content (Resende et al. 2019; Garimella and Eckles 2020) which friends and family may forward through messages, images, or videos. With lower overall usage, however, participants are less likely to engage with such forwarded content in the first place. Third, simply raising the barriers to passive consumption may change the type of information people consume: turning off automatic multimedia downloads—effectively introducing barriers to the first visualization of media content on WhatsApp—requires users to make an active choice to view content, thereby increasing their likelihood of both paying attention to (1) whether they wish to view the content and (2) the content's information itself. This small adjustment in how WhatsApp users interact with the platform could substantially reduce the passive intake of misinformation.

In short, both raising the stakes for multimedia consumption and encouraging selective engagement through overall consumption reduction can create an environment where misinformation struggles to gain a foothold, leading to a more informed and discerning user base. Yet, prioritizing certain forms of engagement with social media may also reduce the consumption of *truthful* news published during the electoral cycle. Since participants are making determinations about what to consume based on sources (close networks versus larger channels) and attributes (multimedia), it may be difficult to otherwise differentiate content types. Further, news consumption on social media platforms tends to be incidental through large groups, meaning that people do not prioritize seeking out news directly on a daily basis (Boczkowski, Mitchelstein and Matassi 2018; Masip et al. 2021). Insofar as large channels and broader networks expose participants to both true and false news, selection out of these networks can reduce true news consumption; individuals may prefer to limit their usage to social interactions instead of looking at news content.

H3a: Reducing WhatsApp usage reduces exposure to *false* information.
H3b: Reducing WhatsApp usage reduces exposure to *true* news.

2.2 Consequences For Information Belief

By reshaping information exposure, reducing WhatsApp usage ought to shift how individuals form and solidify their beliefs. Existing research has demonstrated that repeated exposure to information increases the likelihood of believing it is true, regardless of accuracy (Pennycook, Cannon and Rand 2018). For example, once misinformation is consumed, false beliefs may be difficult to correct due to motivated reasoning (Ecker et al. 2022; Martel, Pennycook and Rand 2020). This finding underscores the importance of consumption patterns in shaping public opinion: when individuals are exposed to less content, they might be more skeptical and less confident in the veracity of the information they encounter. This reduced certainty impacts the acceptance of both false and true information, making it harder for misinformation to take root but also introducing doubt in truthful news circulating on social media. We therefore hypothesize the following:

H4a: Reducing WhatsApp usage increases the likelihood of identifying *false* information as false.
H4b: Reducing WhatsApp usage decreases the likelihood of identifying *true* information as true.

2.3 Consequences For Political Polarization

Political polarization is often cited as a negative consequence of social media usage and broader internet access (Allcott et al. 2020; Arugute, Calvo and Ventura 2022; Lelkes, Sood and Iyengar 2017; Settle 2018). When societies are deeply divided along social or ideological lines, and when politics becomes factionalized, policy disagreements can escalate into negative perceptions of the opposition and social alienation of outgroup members (Iyengar et al. 2019). This, in turn, can weaken social cohesion, erode democratic norms, exacerbate policy gridlock, and even fuel political violence (Kingzette et al. 2021; Svolik 2019; Piazza 2023; Badrinathan, Chauchard and Siddiqui 2024).

We theorize that reducing WhatsApp usage could minimize these dynamics by filtering individuals out of low-quality online spaces where polarizing content is most prevalent. Polarization does not stem from (mis)information in a vacuum, but from the wider context in which it circulates. This context, characterized by toxic discourse and echo-chamber dynamics (Barberá 2020), repeatedly exposes individuals to extreme views and selective amplification of content at a high volume (Aruguete, Calvo and Ventura 2023). By disengaging from such spaces and from viral polarizing content circulating on WhatsApp, individuals may distance themselves from polarizing interactions and minimize the formation of associated beliefs.

Thus, reducing WhatsApp usage may mitigate social divisions, presented as either affective polarization, where political partisans view opposing groups with hostility (Iyengar, Sood and Lelkes 2012), or identity-based prejudice rooted in ethnic or racial divisions (Wilkinson 2006). During elections, (mis)information about outgroups is pervasive (Rathje, Van Bavel and Van Der Linden 2021), which can exacerbate social and political hostility (Kingzette et al. 2021; Jenke 2024). In addition, politically rooted misinformation and misperceptions about outgroups play a key role in heightening polarization (Druckman et al. 2023; Voelkel et al. 2023). Further, when exposed to contentious and uncivil arguments—common to group chat dynamics on WhatsApp (Chauchard and Garimella 2022)—voters tend to double-down on their preferences and use increasingly toxic speech (Kim et al. 2021), which can exacerbate attitude polarization (Velez and Liu 2024). Building on evidence that reducing exposure to misinformation and polarizing content can mitigate affective polarization (Druckman et al. 2023; Levy 2021), we hypothesize:

H5: Reducing WhatsApp usage reduces affective partisan polarization.
H6: Reducing WhatsApp usage reduces identity-based outgroup prejudice.

Reducing WhatsApp usage may also reduce ideological divisions, particularly in the Global South context, where social media has become the primary way through which many voters report receiving news and learning about elections (Newman et al. 2024), and where political elites rely heavily on it to communicate with their constituents (Bessone et al. 2022; Wirtschafter et al. 2024). We consider issue polarization, where people become increasingly divided over specific political topics and party platforms, and candidate favorability, where people more strongly prefer the inparty candidate versus candidates from other parties. Decreased engagement with negative political discourse, particularly in relation to ongoing discussions about policy proposals and party campaigns, could reduce the likelihood that people are polarized by partisan media coverage, which in turn might lead them to better consider issues themselves (Velez and Liu 2024). More mechanically, through reduced exposure to *true* news, people may also be less certain about the issues altogether. Similarly, reducing exposure to biased or inflammatory content during the election period ought to reduce candidate favorability: by seeing less ingroup versus outgroup rhetoric, people ought to be less likely to intensely favor their preferred party's candidate(s) and oppose outparty candidate(s). We therefore hypothesize:

H7: Reducing WhatsApp usage reduces issue polarization.
H8: Reducing WhatsApp usage reduces candidate favorability.

2.4 Substitution Effects and Well-Being

Finally, reducing WhatsApp usage may generate important non-political effects if participants choose to prioritize communication with close friends and family and reduce time spent in low-quality, impersonal spaces such as large groups or channels. We expect this shift to increase subjective well-being and to free up time for other activities. Prior research highlights several potential mechanisms for how social media usage might reduce subjective well-being—such as reduced face-to-face interaction, increased social comparison (Kross et al. 2021; Twenge and Campbell 2018), and information overload (Matthes et al. 2020; Goyanes, Ardèvol-Abreu and Gil de Zúñiga 2023). We therefore expect that reductions in WhatsApp usage, by enhancing the quality of online interactions and creating opportunities for offline experiences, should jointly improve participant well-being.[3]

3 Experimental Design

In 2024, we implemented our field experiment preceding elections in three major Global South democracies: India, South Africa, and Brazil. India and South Africa held general elections in the spring—India's polling ran from April 19 and June 1 with results announced on June 4, and South Africa's election was held on May 29 with results announced on June 2. Our four-week usage reduction experiment ran concurrently (from April 29 to May 27) in these two countries. We then repeated our experiment in Brazil (from September 9 to October 4) ahead of its municipal elections on October 6. We randomly assigned half of the treated participants to change their WhatsApp usage, either by reducing multimedia consumption (as in Ventura et al. 2025) or reducing total time spent on the app during the four weeks. We then administered a final survey to measure the effects of participating in the experiment.[4] We discuss each stage of the study in further detail below.

3.1 Recruitment

We recruited participants using Meta Advertisements, posting ads on Facebook, Instagram, and Messenger (see Appendix Section A.1 for examples). The ads directed participants to a short Qualtrics survey where they answered questions pertinent to study eligibility and block randomization. These include questions about respondents' demographics, their social media habits (particularly focusing on WhatsApp usage), and their willingness to join a four-week study that potentially involves reducing WhatsApp usage. In total, we recruited 1,310 eligible participants in India, 2,884 in South Africa, and 2,067 in Brazil.

[3] We note that these expectations and associated analyses are exploratory, as they were pre-registered as additional research questions without directional hypotheses.

[4] All of our surveys carried monetary incentives. Appendix Section A.3 details our incentive structure, which was communicated to participants during recruitment.

3.2 Baseline & Treatment Assignment

We then invited all eligible participants to join our study Participants were given a baseline survey, which had three components. First, we collected pre-treatment covariates of interest and pre-treatment measurements of a subset of our eventual outcomes. Second, participants provided their baseline WhatsApp usage information by submitting screenshots of their WhatsApp settings page (explained further below). Third, at the end of the survey, we informed participants of their pre-randomized treatment condition. Participants were block-randomized on age, education, gender and self-reported WhatsApp usage, and then assigned to one of four equal-sized groups (see Table 1). In total, 678 (52%) individuals in India, 820 (28.5%) in South Africa, and 928 (44%) in Brazil successfully enrolled in the experiment.[5] In Appendix Section C.2, we test for baseline differences in pre-treatment covariates between participants invited and participants enrolled. Overall, participants who enrolled were younger, more educated, more digitally savvy, and heavier WhatsApp users than the wider sample of eligible participants.[6]

To facilitate compliance, we impose barriers to meaningfully reduce usage. Participants assigned to *Multimedia* were asked to turn off their automatic download of multimedia (images, videos, documents, and audio). This blurred the content they received, which they could tap to download and view individually. Participants assigned to *Time* were asked to set their WhatsApp app timer to 10 minutes per day. The timer reminded them when they passed their screen time daily limit, but allowed them to override the option for a few more minutes. Thus, neither of these interventions imposed hard constraints on participants' WhatsApp experience, but rather added friction to participants' regular WhatsApp usage.

To avoid one potential source of differential attrition, our experimental design only informed participants of their treatment assignment after they uploaded their first screenshot showing their

[5] Most people contacted via email and WhatsApp did not start the baseline survey and thus never saw their treatment assignment. The share of people who successfully enrolled in the experiment *conditional* on having seen treatment assignment is 83.3% (n=1,111) in Brazil, 90% (n=755) in India, and 77% (n=1,056) in South Africa.

[6] We discuss these differences in detail, including implications for inference, in Appendix Section C.3.

baseline time or media usage. To ensure that all participants were willing and able to join the experiment, we also asked the control group to spend three days performing the same task asked of the treatment group—which in practice meant a separate control group for each treatment. Thus, control group participants were nominally "treated" for a brief period; however, we do not expect effects from this short, suggestive intervention to persist until the endline survey four weeks later. Table 1 presents a summary of the instructions for each group.

Table 1: Summary of Instructions for Treatment and Control Groups

Condition	**Instructions & Steps**
Multimedia	i) Submit a screenshot showing their WhatsApp media storage statistics (Appendix Figure A6). ii) Informed of treatment: Do not consume any multimedia content on WhatsApp for four weeks. iii) Given instructions on how to turn off automatic media downloads on WhatsApp. iv) Submit a screenshot showing that automatic media downloads has been turned off (Appendix Figure A7).
Control (Multimedia)	i) Submit a screenshot showing their WhatsApp media storage statistics (Appendix Figure A6). ii) Informed of treatment: Do not consume any multimedia content on WhatsApp for three days.
Time	i) Submit a screenshot showing their WhatsApp screen time (Appendix Figure A6). ii) Informed of treatment: Limit WhatsApp usage to 10 minutes per day for four weeks. iii) Given instructions on how to set a 10-minute daily timer for WhatsApp. iv) Submit a screenshot showing that the timer has been set (Appendix Figure A7).
Control (Time)	i) Submit a screenshot showing their WhatsApp screen time (Appendix Figure A6). ii) Informed of treatment: Limit WhatsApp usage to 10 minutes per day for three days.

3.3 Compliance Tasks

We monitored compliance by asking respondents to provide us with screenshots from their mobile devices showing their WhatsApp screen time or media consumption statistics, depending on their treatment assignment. Once a week, we sent participants a short survey prompting them to upload these screenshots. In total, participants submitted four such screenshots, in addition to the initial screenshot collected during the baseline survey.

Participants in the *Time* conditions were asked to send screenshots of their daily WhatsApp usage (in minutes) for every week of the experiment. This information is available in the Settings app of standard mobile devices. Participants in the *Multimedia* conditions were asked to send screenshots of their WhatsApp storage information. This page is available within WhatsApp and records, in bytes, the volume of media received via WhatsApp. Appendix Figure A6 contains examples of these screenshots.

3.4 Post-Treatment Survey

After the four-week experiment, we invited respondents to a final survey. The outcomes collected through this survey are described in detail in the next section, where we present our findings regarding the effects of reducing WhatsApp usage during elections.[7] The survey was sent to Indian and South African participants on May 27, a few days before election results were announced in their countries, and to Brazilian participants on October 4, three days before the election. We chose to send the surveys before the announcement of results in India and South Africa in order to avoid the potential risks of restricting participants' access to WhatsApp post-election. In Brazil, the results are announced hours after polls are closed, leading us to send the survey before the voting day. In total, 653 (96%) people completed the post-treatment survey in India, 742 (90%) in South Africa, and 825 (89%) in Brazil.

We investigate differential attrition in Appendix Section C.3. Combining data from all three countries, 2,220 out of 2,425 enrolled participants completed the post-treatment survey, yielding an attrition rate of 8.4%. There is no evidence of differential attrition between treatment (8.5%) and control (8.3%). We further examine the presence of selective attrition by comparing baseline char-

[7] See Appendix Section B for a summarized presentation.

acteristics in pre-treatment covariates and outcomes between attriters and those who completed the study across treatment and control (Ghanem, Hirshleifer and Ortiz-Becerra 2023); we find no evidence of this (Appendix Table C10).

3.5 Estimation

In our regression analyses, we estimate the adjusted intention-to-treat (ITT) effect of the corresponding treatment on our outcomes of interest using OLS estimators.[8] As pre-registered, we add as covariates the variables used in our block randomization procedure (age, gender, education, and self-reported WhatsApp usage) and additional pre-treatment variables selected via Lasso for each outcome. Our primary specification pools the two treatment arms across all countries; see Appendix Section D.1 for unpooled treatment arm effects.[9] For each outcome, we also present pooled treatment effects by country and unpooled country-level effects. To account for the distinct baselines of the outcomes in each country, all models use a multilevel estimation with random intercepts at the country level. All confidence intervals use a two-sided test with $p < 0.05$ as our measure of statistical significance.[10]

4 Results

We present the following results. First, we assess the extent to which treated participants actually reduced WhatsApp usage relative to the control group (Section 4.1). We then present the results of our pre-registered analyses relating to information consumption (Section 4.2), belief accuracy (Section 4.3), broader social and political attitudes (Section 4.4), and non-political outcomes (Section 4.5). Lastly, we consider heterogeneous effects at the participant and country level to discuss

[8] Appendix Section F presents unadjusted ITT results. Appendix Section D.7 presents CACE models.

[9] In the unpooled models, to increase statistical power, we compare each treatment arm with a pooled control group. Appendix Section D.8 shows that all primary outcomes have non-distinguishable mean differences across the media and time control groups.

[10] Appendix Section D.5 presents results with multiple hypothesis adjustments.

how consumption patterns shape outcomes (Section 4.6).

4.1 Assessing Compliance

At the end of every week of the experiment period, to examine whether our intervention reduced WhatsApp usage and consumption as intended, we collected a screenshot showing one's weekly WhatsApp screen time from participants in the *Time* arm and a screenshot showing one's cumulative WhatsApp data consumption in the *Multimedia* arm.

We find that treated users logged substantially lower levels of WhatsApp screen time and bytes downloaded compared to control group users. In the *Time* arm, pooling across countries, only 20% used WhatsApp for less than 10 minutes at baseline, compared to 65% in the treatment condition during the experiment. Meanwhile, 77% of the treated users in the *Multimedia* arm consumed fewer megabytes than the average control group user in their country. Thus, our treatment assignment worked as intended in creating two distinct groups with substantially different levels of WhatsApp usage during elections.[11]

While treatment assignment significantly reduced usage in all three countries, we do observe considerable variation in control group usage levels across countries, with our Brazilian respondents recording significantly higher baseline screen time and media consumption. We also observe that there is a stronger degree of compliance in the *Multimedia* arm compared to the *Time* arm, suggesting that, given the same monetary incentives, duration of treatment, and baseline usage levels, individuals are more willing and able to stop consuming multimedia content on WhatsApp compared to limiting WhatsApp usage to 10 minutes per day. This is an important descriptive finding for future academic research and policymaking on WhatsApp, confirming our priors that a full WhatsApp deactivation is infeasible in settings where it is much more embedded

[11] If a user submitted a screenshot that was doctored, a duplicate of a previous week's submission, from a different phone as their previous submissions, or indicated their data consumption statistics had been manually reset during the experiment period, we classify them as having violated study rules and mark their screenshots as missing. About 4% of our sample either violated one of these rules or simply did not submit screenshots; reassuringly, country or treatment status does not predict violations or missingness.

in users' lives relative to social media platforms like Facebook would be. Appendix Section D.7 discusses compliance estimators in further detail.

4.2 The Effects of WhatsApp Usage on Information Consumption

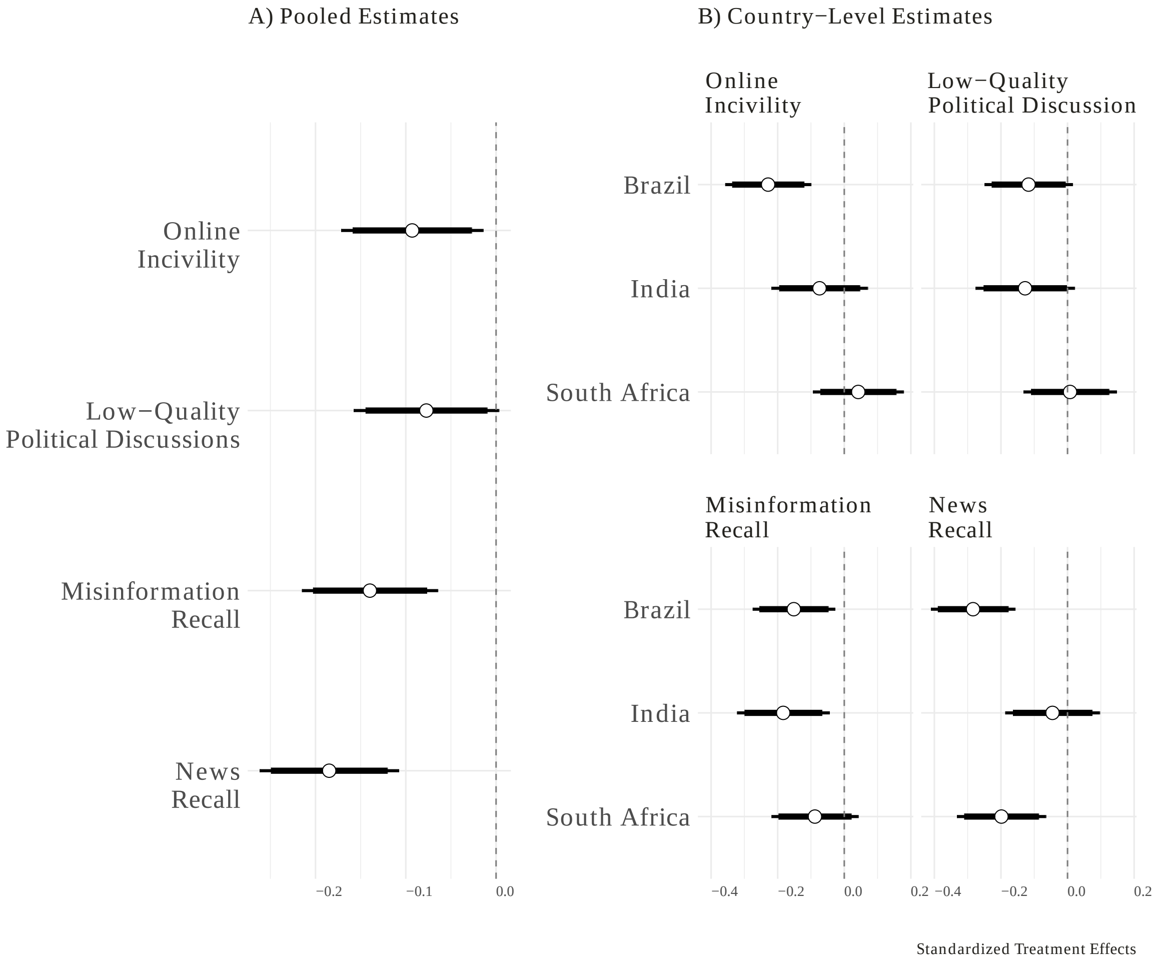
4.2.1 Online Incivility and Quality of Political Discussions

Figure 1 presents the effects of reducing WhatsApp usage on our outcomes related to quality of information consumption: exposure to toxic content online and to low-quality political discussions online and offline. For the first outcome, we ask them if, during the past month, they received hostile comments online or saw online comments that were rude or disrespectful. We build an index with these two measures called *Online Incivility*. For the second outcome, we use a battery of items asking participants about their exposure to discussions that made them angry (Allcott et al. 2020), that they consider uncivil (Rossini 2022; Chen 2017), and that helped them reduce prejudice towards outgroups (Allcott et al. 2020). We build an index with these three measures called *Low-Quality Political Discussions*.

Figure 1 shows that reducing WhatsApp usage produced a significant decrease in self-reported exposure to online toxic speech and to low-quality discussions about politics, consistent with **H1** and **H2**. The intention-to-treat analysis shows a reduction of 0.09 SD (p-value < 0.01) in exposure to online toxicity and of 0.07 SD (p-value < 0.10) in exposure to low-quality political discussion. For online toxicity, the effects are primarily driven by the Brazilian sample, while for low-quality political discussion, both Brazil and India show consistent reductions. The treatment's particular effectiveness on these outcomes in Brazil reflect a combination of two factors: (1) Brazilian participants' somewhat higher baseline exposure to incivility online (see Appendix Figure D18), which were reduced as a result of treatment, and (2) their greater tendency to substitute WhatsApp usage with offline activities such as hobbies or time with friends (see Figure 4). This difference highlights how variations in the informational environment influence the effects of social media reduction interventions.

In Appendix Figure D8, we further disaggregated our two treatment arms to explore their differential effects on these outcomes. We find that reductions in exposure to online incivility are primarily driven by turning off automatic media downloads, whereas reduced exposure to low-quality political discussions are predominantly driven by capping overall WhatsApp screen time. It is possible that reduced exposure to inflammatory media content mitigates perceptions of online exposure to uncivil content, while minimal engagement with the platform as a whole limits participation in toxic political discussions. Generally, these effects echo anecdotal and descriptive evidence of the role of social media in increasing exposure to hostile content and low-quality information about political issues (Bor and Petersen 2022; Recuero, Soares and Vinhas 2021).

Figure 1: Treatment Effects on Information Consumption



Notes: See Appendix Tables G23 and G24 for full regression results.

4.2.2 Misinformation and News Exposure

We hypothesized that reducing WhatsApp usage ought to reduce exposure to misinformation and true news. In our endline survey, we provided participants with two distinct sets of headlines—news headlines and misinformation—to assess this set of hypotheses. To measure *misinformation exposure*, we used major fact-checking websites in each country to identify four false stories that circulated widely on social media during the election. We selected a diverse set of misinformation headlines corresponding to different topics, political leanings, and weeks of the experimental period. To measure *true news exposure*, we selected six stories that appeared on major news organizations' websites in each country during the same period.[12] Four were presented verbatim, while two were altered to create placebo false news by reversing a key fact from the original headline, helping to guard against acquiescence bias.[13] The full text of each headline is listed in Appendix Table B4.

To construct our exposure measures, we asked participants whether they had seen each headline in the past 30 days. `Misinformation Recall` counts the number of misinformation rumors (0-4) that a participant reported seeing; similarly, `News Recall` counts the number of true news stories (0-4) recalled.[14]

Figure 1 shows that, consistent with **H3a**, limiting WhatsApp usage—be it by maintaining time limits or by refraining from consuming multimedia content— reduced recall of political misinformation circulating widely during the intervention. The intention-to-treat analysis shows a reduction in exposure to misinformation of 0.14 SD (p-value < 0.01). In support of **H3b**, we also find a 0.19 SD (p-value < 0.01) decrease in recall of true news stories. In Appendix Figure D8,

[12] To identify salient news headlines, we scraped Google News daily throughout the four-week experiment period and selected a representative sample of six stories covering different issue areas, partisan groups, and time points in the experiment.

[13] This procedure follows prior deactivation studies (Allcott et al. 2020, 2024; Asimovic et al. 2021; Asimovic, Nagler and Tucker 2023; Arceneaux et al. 2023; Ventura et al. 2025).

[14] We drop the two placebo headlines in our analyses of information exposure, since these were headlines that were artificially modified by us and which, by definition, respondents could not have seen.

we present unpooled treatment effects and show that the two types of WhatsApp usage reduction decreased exposure to misinformation and true news to a similar degree.

While the overall effect sizes are comparable across misinformation stories and true news stories, we note from Figure 1(B) that the reduction in misinformation recall is consistent across the three countries, while the decline in true news recall is observed predominantly in South Africa and Brazil. The different finding in India may be driven by different news consumption patterns: in Appendix Figure D19, we find that baseline news exposure in India is higher than in South Africa and Brazil, and thus participants may have more outlets through which they were able to stay informed about current events.

Notably, we also asked participants to recall how much true and false news they saw on social media over the past month. In Appendix Figure D11, we show that participants reported seeing less false news on social media but not true news. In other words, although actual consumption of both types of information declined due to treatment, treated participants were only cognizant of reductions in misinformation exposure. This asymmetry suggests that individuals may be less attuned to their consumption of true news on social media more broadly.

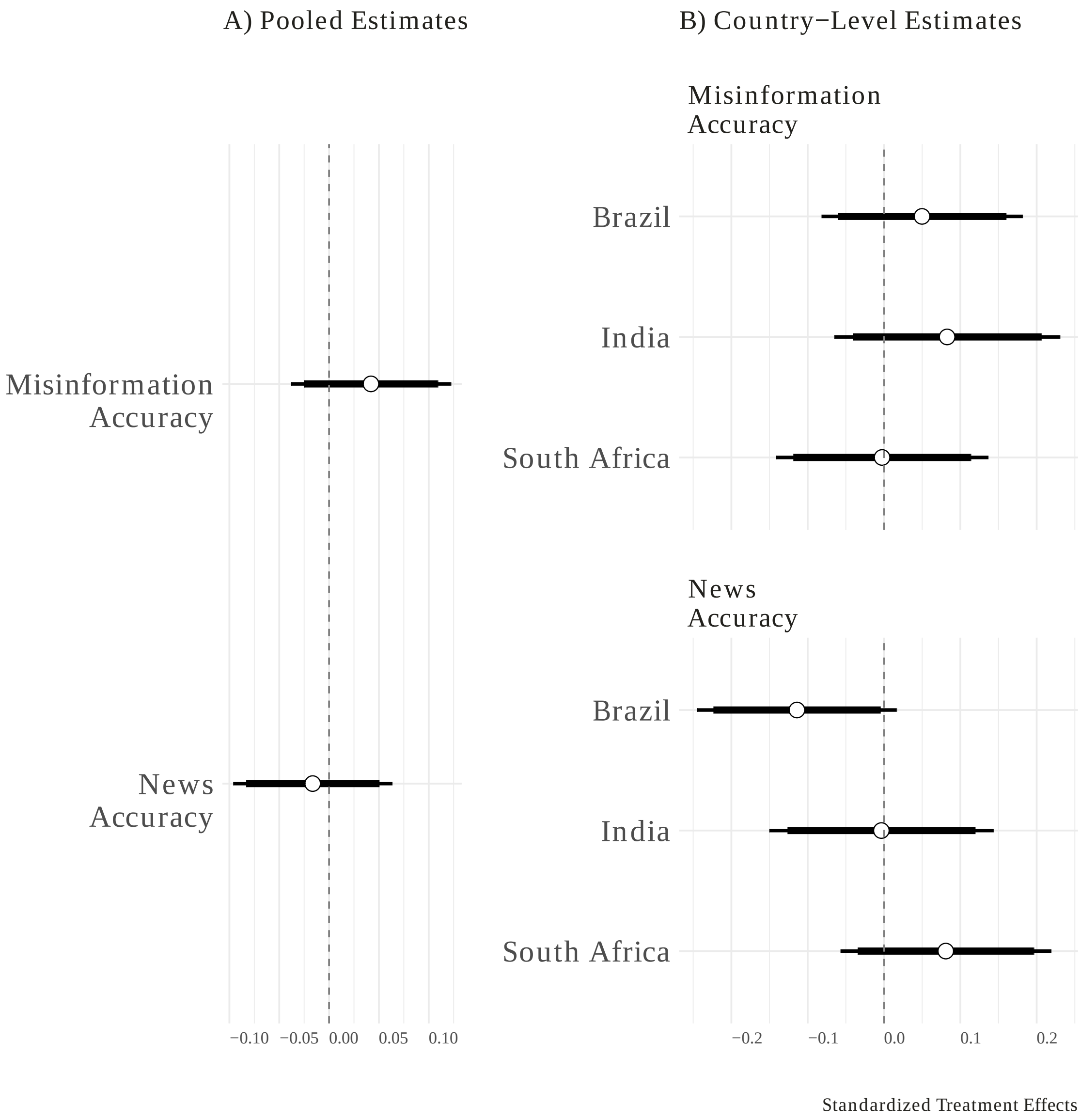
4.3 Downstream Consequences For Belief Accuracy

Our results so far show strong shifts in information consumption. We next examine how these changes affect belief accuracy. After eliciting recall, we asked participants the extent to which they think each headline is accurate. We use these responses to construct two measures of belief accuracy by counting the number of headlines correctly identified as false or true: `Misinformation Accuracy`, summing the number of misinformation headlines correctly identified as false (0-4), and `News Accuracy`, summing the number of news stories and placebo fakes accurately identified as true and false, respectively (0-6).

Figure 2(A) presents pooled estimates corresponding to belief in misinformation stories and knowledge of true news stories. We find limited support for **H4a** and **H4b**. Although the treatments significantly reduced exposure to both true and false headlines, they do not have a statistically significant effect on belief accuracy for either. These null effects align with prior deactivation studies, which have also not shifted accuracy perceptions of false information (Allcott et al. 2020;

Ventura et al. 2025) or general news knowledge (Allcott et al. 2024). Indeed, changes in (self-reported) exposure do not directly translate to changes in belief accuracy, suggesting a potentially more complex mapping from exposure to belief formation than posited in extant literature.

Figure 2: Treatment Effects on Belief Accuracy



Notes: See Appendix Tables G23 and G24 for full regression results.

4.4 Downstream Consequences on Political Attitudes

We theorized that reducing WhatsApp usage ought to also have downstream consequences for users' political attitudes by limiting exposure to negative discourse and (mis)information. In Figure 3, we examine our treatments' effects on partisan polarization (**H5**), identity-based outgroup prejudice (**H6)**, issue polarization (**H7**), and candidate favorability (**H8)**.

4.4.1 Political Polarization and Identity-Based Prejudice

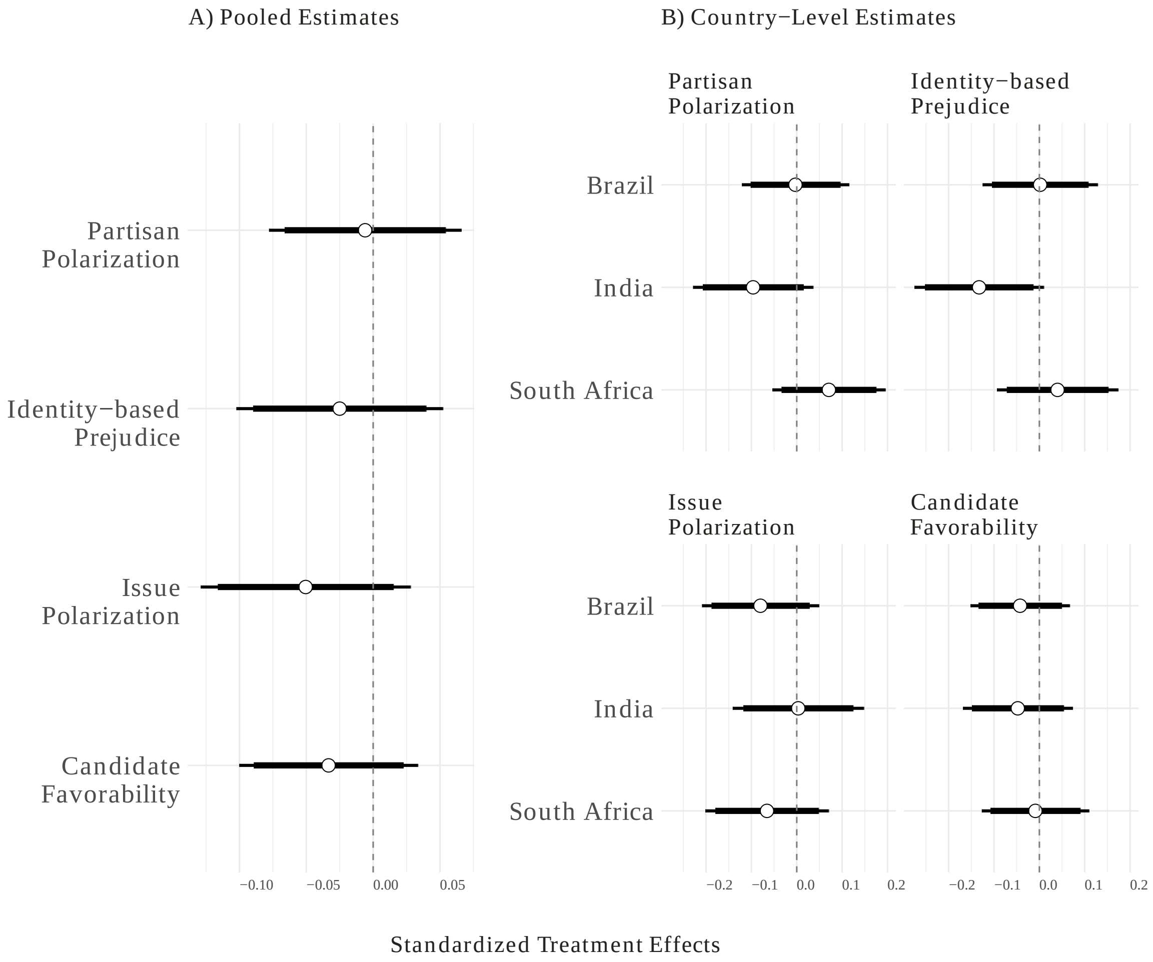
To explore effects on partisan polarization (**H5**), we collected participants' views about voters of the two largest parties in their country, which also represent the main opposing coalitions in the 2024 elections. These were the Bharatiya Janata Party (BJP) and the Indian National Congress (Congress Party) in India, the African National Congress (ANC) and the Democratic Alliance (DA) in South Africa, and the Partido Liberal (PL) and the Partido dos Trabalhadores (PT) in Brazil. We define each participant's *in-party* as the party representing the coalition they would prefer to see win the election, and their *out-party* as the party leading the other coalition. In India, 68.7% (31.3%) of our sample listed the BJP-led coalition (Congress-led coalition) as their preferred choice. In South Africa, 49.1% (50.9%) preferred the ANC-led coalition (DA-led coalition). In Brazil, 44.2% (55.8%) preferred the PL-led coalition (PT-led coalition).

We use three main outcomes to measure partisan polarization. First, we capture overall feelings towards the two parties' voters using a 7-point feelings scale. Second, we present a list of social interactions—such as watching a sports game together or living in the same neighborhood—and ask participants which they would be willing to do with voters of each party. Third, we asked participants to read a list of positive and negative traits and indicate which might describe a typical voter of each party. For each of the three items, we calculate the absolute distance between participants' views about these two parties' supporters. We then aggregate these measures into our `Partisan Polarization` index (see Appendix Table B2).

In Figure 3(A), we find no support for **H5**: while our treatments did reduce exposure to partisan misinformation, online incivility, and low-quality political content, these changes in the informational environment did not mitigate partisan animosity and polarization.

Similarly, **H6** predicted a reduction in identity-based (ethnic or racial) out-group prejudice fol-

Figure 3: Treatment Effects on Political Attitudes



Notes: See Appendix Tables G23 and G24 for full regression results.

lowing our detected changes in the participants' informational environment. Political and societal dynamics in India and South Africa are shaped by identity-based prejudice in similar ways, and thus our analysis explores whether reducing WhatsApp usage may reduce intergroup animosity in these two countries.[15] We focus on measuring post-treatment attitudes about the two most politically salient groups in each country: Hindus and Muslims in India and Blacks and Whites

[15] This analysis focuses only on India and South Africa, as these issues are less central to contemporary Brazilian politics. However, as a validity check, we included and assessed one measure of racial prejudice in our Brazil surveys.

in South Africa.[16] In India, there is a history of conflict between Hindus, the dominant religious majority, and Muslims, the largest religious minority. In recent years, WhatsApp has been used by political elites to spread misinformation and harmful content that demonizes Muslims (Chauchard and Garimella 2022; Saha et al. 2021). In South Africa, we examine prejudice along the main racial cleavage (i.e., Black and White South Africans). Legacies of apartheid in South Africa has kept race salient in politics and a frequent subject of misinformation (Allen 2021; Bowles et al. 2025).

Mirroring our partisan polarization estimation, our `Identity-based Prejudice` outcome comprises the same three items on overall feelings towards, willingness to engage in social activities with, and typical traits associated with, members of different groups. Our prejudice index aggregates responses to these three items, where higher values represent a larger gap between responses about the two groups (see Appendix Table B3). In Brazil, we only included the 7-point feelings scale, asking respondents to rate their feelings about White and Black Brazilians.

The second outcome in Figure 3(A) presents treatment effects on identity-based prejudice.[17] Pooling across countries and treatment types yields null results, but obscures an important within-country finding. While we do not uncover treatment effects in South Africa and Brazil, Figure 3(B) shows that there is a marginally significant reduction in ethnic prejudice in India for participants assigned to the time treatment ($d = -0.13$, p-value $= 0.067$). Given journalistic and scholarly accounts of substantial anti-Muslim content circulating on WhatsApp in India (Chauchard and Garimella 2022; Saha et al. 2021), especially during elections, we surmise that a drastic reduction in exposure to such content somewhat mitigates anti-Muslim bias among treated users, relative to the control group that presumably sees content that portrays Muslims negatively.

[16] Although many ethnic and racial groups exist in these countries, most WhatsApp discourse focuses on Hindus/Muslims in India and Blacks/Whites in South Africa. We therefore do not expect WhatsApp usage reduction to affect attitudes about other groups. To mitigate survey fatigue and priming, we limit questions to these main groups.

[17] It should be noted that our samples are quite skewed in both countries, which is to be expected given the composition of the population in each country. In India, 78.6% of our sample are Hindu. In South Africa, 81.3% are Black. The results are therefore predominantly driven by Hindus in India and Blacks in South Africa.

4.4.2 Issue Polarization and Candidate Favorability

We also explore how reducing WhatsApp use influenced attitudes towards election-related issues and shaped evaluations of political candidates. To measure issue polarization (**H7**), we elicit opinions about six issues that drew public attention during the election period in each country, covering democratic norms, intergroup relations, immigration policy, and news events. Our `Issue Polarization` index is constructed by summing the standardized absolute differences between each participant's agreement level and the control group's mean agreement for each issue.[18] Positive values indicate stronger issue polarization, with the participant moving away from the average baseline issue agreement, and negative values mean the participant moved toward the average baseline issue agreement.

To measure candidate favorability (**H8**), we use a standard feelings scale featuring major political leaders in each country, including the executive branch incumbent.[19] Our measure is intended to capture whether WhatsApp usage makes voters more favorable toward their preferred party's candidates and less favorable towards alternatives. Our `Candidate Favorability` index takes the absolute difference between the favorability ratings of participants' main political ingroup candidate and the average favorability towards candidates from the alternative parties.

The third and fourth outcomes in Figure 3(A) present the pooled effects of our intervention on issue polarization and candidate favorability respectively. Overall, the effects of reducing WhatsApp usage on both outcomes are in the hypothesized direction but are not statistically sig-

[18] This measure adapts the issue agreement approach from Allcott et al. (2020). The original measure uses partisan groups as reference points. We depart from this practice because political issues in these countries often lack systematic partisan cleavages. For example, in India and Brazil, only three of six issues align clearly with partisan divisions; in South Africa, none do. We therefore use average issue agreement in the control group as our reference point instead of partisan benchmarks.

[19] In India, we ask about Narendra Modi (BJP), Mallikarjun Kharge (Congress), Rahul Gandhi (Congress), and Yogi Adityanath (BJP). In South Africa, we ask about Cyril Ramaphosa (ANC), John Steenhuisen (DA), Julius Malema (EFF), and Jacob Zuma (MK). In Brazil, we ask about Lula (PT) and Jair Bolsonaro (PL).

nificant at conventional levels.[20]

In sum, the null effects across all political measures suggest that reducing WhatsApp usage may not be enough to shift broader political behaviors and attitudes, *despite* altering the information environment and exposure. These findings align with theories of media minimal effects and are consistent with previous deactivation studies and other social media field experiments deployed in election periods (Arceneaux et al. 2023; Allcott et al. 2024; Nyhan et al. 2023; Guess et al. 2023; Ventura et al. 2025).

4.5 Non-Political Downstream Effects

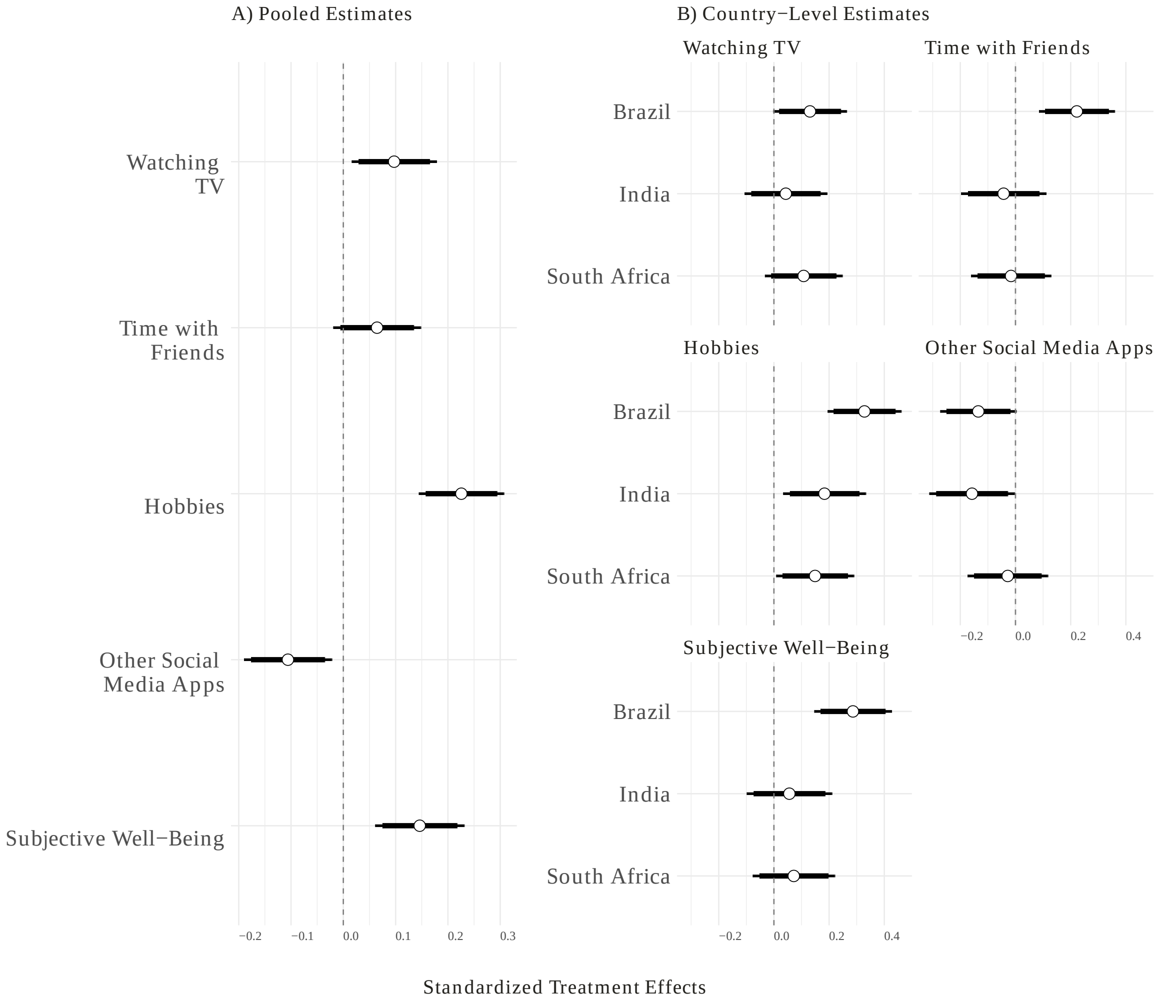
While reducing WhatsApp usage did not shift political attitudes, our treatments may have elicited other changes at the individual level that were not explicitly political in nature. We consider (1) substitution away from WhatsApp to different activities and (2) changes in subjective well-being.

Figure 4 presents substitution effects. We find that our incentives to reduce WhatsApp usage pushed individuals to substitute their time with offline activities. For instance, we detect an increase of 0.09 SD (p-value $<$ 0.01) in watching TV, 0.06 SD (p-value $=$ 0.13) in participating in social activities with friends, and 0.22 SD (p-value $<$ 0.01) in spending time on other hobbies. Meanwhile, reducing WhatsApp usage also led participants to reducing their (self-reported) usage of other social media apps by 0.10 SD (p-value $<$ 0.01). However, these effects are not concentrated in a specific platform. In Appendix Figure D13, we find null results using an alternative set of questions where we asked for usage across specific social media platforms including Telegram, Facebook, TikTok, Twitter, and others. This suggests that participants reduced their WhatsApp usage and may have extrapolated this reduction to their overall social media time. These substitution effects are strongest in Brazil, consistent with its higher compliance rates, but are in a similar direction in India and South Africa as well.

On subjective well-being, participants were asked to rate how often they felt (a) anxious, (b) depressed, (c) satisfied with life, (d) happy with their appearance, and (e) isolated from family in the past few weeks on a five-point scale ranging from ”Never” to ”All the time.” We aggregate

[20] In Appendix D.4, we examine downstream consequences on respondents’ level of political interest and turnout; we show that our treatments did not shift either.

Figure 4: Non-Political Downstream Effects of Reducing WhatsApp Usage



Notes: See Appendix Tables G27 and G28 for full regression results.

these responses into a `Subjective Well-Being` index.[21] This constitutes the final outcome in Figure 4. Overall, reducing WhatsApp usage did improve participants' subjective well-being: compared to those in the control condition, treated participants reported an 0.14 SD increase in subjective well-being (p-value < 0.01).

When disaggregating the results in Figure 3(B), we note that this effect is larger and more precisely estimated in Brazil. These changes occur alongside more substantial substitution effects

[21] Negative indicators (anxiety, depression, and isolation) are reverse-coded in the index such that higher values correspond to improvements in well-being.

from social media to offline activities in Brazil. This country-level difference suggests that social media reduction interventions by themselves might not produce direct improvements in well-being if they are not associated with other pro-social offline activities and more time engaging in direct, in-person interactions. These differences emphasize the critical gains from our multicountry design, as they highlight how participant behaviors in each country condition the effects of social media usage.

4.6 How Consumption Patterns Shape Outcomes across Countries and Individuals

To further unpack our main findings, we consider how individual consumption patterns and different country contexts shaped the observed effects.

Conducting pre-registered analyses of heterogeneous treatment effects, we find that WhatsApp usage for news and politics produced weak moderation effects, while age and digital literacy yield null results. Instead, the intervention was primarily effective among participants who used WhatsApp heavily at baseline (see Appendix Figure D14). For these participants, who reported using WhatsApp more than the median user in our sample did, we also observe greater reductions in exposure to online hostility and low-quality political discussions as well as in recall of misinformation and true news headlines. Most interestingly, we observe improvements in their ability to accurately identify misinformation rumors, suggesting that reduced exposure to misinformation can improve skepticism of false content among high-dosage WhatsApp users without shifting confidence in true news.

Differences by the strength of partisanship further highlight the role of consumption choices (Appendix Figure D15). Participants with weaker partisan identities showed stronger reductions in exposure to low-quality and hostile content, as well as greater improvements in information recall, in comparison to participants with stronger partisan identities. Weak partisans therefore appeared more willing to disengage from toxic or politicized spaces when forced to limit their WhatsApp usage. In contrast, strong partisans exhibited smaller informational changes, suggesting that they were more likely to remain active in highly partisan spaces even when instructed to reduce WhatsApp use. Still, the intervention did reduce overall misinformation exposure even among strong partisans, indicating that reducing WhatsApp usage did provide a buffer against false content.

Interestingly, reductions in misinformation exposure may have had some downstream consequences on political attitudes, as we find that strong partisans also exhibited reductions in issue polarization. This finding has two implications. First, this pattern may occur because participants who were strongly partisan at baseline were initially both more exposed to and more responsive to partisan cues about salient issues during the election. In other words, among those most engaged with political information prior to our experiment, even modest reductions in (mis)information exposure had noticeable effects on issue-level attitudes. Second, that we find shifts in issue polarization but not other forms of polarization points to a temporal dimension of polarization through WhatsApp exposure: reducing WhatsApp usage may not shift entrenched partisan or affective polarization, which may be more deeply rooted (and particularly so for strong partisans). However, it can reduce issue-specific polarization as elections approach via reduced exposure to misinformation and partisan cues about divisive issues.

Our interpretation of these individual-level patterns are also consistent with cross-country differences in consumption. Reducing WhatsApp usage does not automatically reshape the information environment: it matters whether participants substituted away from the types of content and spaces most associated with negative political outcomes. In Brazil, where large WhatsApp groups are popular and often serve as hubs for political discussion (Kalogeropoulos and Rossini 2025; Newman et al. 2024), reduced WhatsApp usage led to the most substantial declines in exposure to toxicity and measurable gains in subjective well-being. In contrast, in South Africa, our interventions failed to shift exposure to incivility and low-quality political discussions but still reduced (mis)information recall, suggesting that information consumption is woven into daily conversation through interpersonal networks than through explicitly political groups. Finally, at baseline, participants in India were more politically engaged and trusting of social media (Appendix Figure **??**), yet relied less on WhatsApp for news. Correspondingly, reducing WhatsApp usage did not reduce news recall but did reduce misinformation recall.

5 Conclusion

Our four-week online field experiment across three major democracies in the Global South shows that reducing WhatsApp usage limited exposure to uncivil political discourse and misinformation circulating in the weeks before elections. However, consistent with accounts of citizens relying on WhatsApp for true information and news, this reduction in usage also made individuals less likely to recall true news headlines from the election period. Further, these changes in information consumption did not improve belief accuracy, nor did it consistently shift political attitudes, as participants' opinions about partisan groups, ethnic/racial groups, political issues, and political candidates remained stable even as exposure to vitriolic content declined. Our information consumption results are strongest among heavy WhatsApp users—which is consistent with prior work (Aslett et al. 2022; Ventura et al. 2025)—while strong partisans were least likely to alter their usage behavior.

Our findings highlight several trade-offs and limitations. First, interventions that successfully reduce exposure to toxic and false political content may simultaneously limit access to legitimate political information that citizens need for informed democratic participation.[22] Second, individual consumption patterns affect selective exposure, while the social embeddedness of political beliefs constrains the impact of reducing online toxicity and misinformation. Simple reductions in exposure may be insufficient to shift entrenched political attitudes when multiple reinforcing mechanisms across both online and offline environments contribute to their maintenance. Our results point to the need for additional theory-building that accounts for how citizens integrate online and offline sources of political content.

Despite limited political effects, our study highlights the potentially positive *non-political* effects of reducing WhatsApp usage. At the end of the experiment period, participants in our usage reduction treatment groups reported spending more time on their hobbies and with their friends and family; they also exhibited notable improvements in subjective well-being and an overall reduction in social media usage beyond WhatsApp. These results underscore the potential personal and social benefits of interventions that encourage more intentional and limited social media en-

[22] This finding echoes evidence of potential spillover effects from misinformation-reduction interventions (Hoes et al. 2024).

gagement. We anticipate these findings being valuable for policymakers, practitioners, and scholars seeking to identify a balance between the need for robust information flows and the potential harms of unrestrained social media usage.

Participants' reflections on the experience provide a promising foundation when considering the scalability of our intervention. In response to an open-ended question about the experience, many treated participants described it as rewarding and unexpectedly beneficial even if quite challenging at times. Common themes emerged, including the difficulty of avoiding an app as deeply embedded in individuals' lives as WhatsApp is; the benefits of not seeing unnecessary or annoying content due to prioritizing their limited time on the app; and the advantage of gaining significant time to pursue other activities. Notably, 64% of respondents mentioned positive feelings and 52% mentioned negative ones, indicating that many participants candidly acknowledged the pros and cons of such a major lifestyle change. We present a comprehensive analysis of participant perspectives in Appendix E.

Our study's treatment conditions and findings may be instructive for policymakers as they highlight the practical feasibility of reducing WhatsApp usage at scale. Full deactivation is difficult to implement even on platforms like Facebook, and certainly more so on WhatsApp, given its integration into daily social and professional life. Our findings suggest that implementable reductions in WhatsApp usage can meaningfully reduce exposure to misinformation and toxic content without disrupting essential communication. Notably, we observe little difference in outcomes between either reducing *Multimedia* consumption or reducing total *Time*. This is particularly important considering that, while reducing usage to ten minutes per day may be more behaviorally demanding, turning off multimedia downloads is simple. We observed this empirically within our sample: while users in both treatment groups made considerable strides in complying with their assigned treatment, users in the *Multimedia* arm were generally somewhat more successful. These results point to a concrete policy implication, as WhatsApp could make multimedia downloads "off" by default and allow users to opt in when desired. Such a platform-level intervention is minimally disruptive and preserves users' autonomy and access to communication, while also achieving many of the same informational benefits.

Lastly, our results on individual consumption patterns and contextual dependence highlight fruitful avenues for future research. First, by conducting our experiment during the last few weeks of election campaigning—which are characterized by heightened information flows and polarization—we construct “the hardest test” for detecting attitudinal changes. The effects of WhatsApp usage reduction interventions may vary substantially across different moments of the electoral cycle when information environments and users’ receptiveness to new information differ. Future research should explore how similar interventions might function during non-election periods or at earlier stages of electoral cycles when political attitudes might be less crystallized and more amenable to change. Second, we find that the positive effects of reducing social media usage on well-being are context-dependent and further strengthened by substitution patterns; users feel better when reduced social media usage is paired with an increase in offline and face-to-face activities. Future research should consider testing direct social media reduction interventions with incentives to distinct substitution activities in order to better disentangle these effects.

References

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer and Matthew Gentzkow. 2020. "The welfare effects of social media." *American Economic Review* 110(3):629–76.

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

Allen, Karen. 2021. "Social Media, Riots and Consequences." Institute for Security Studies.
URL: https://issafrica.org/iss-today/social-media-riots-and-consequences

Anspach, Nicolas M and Taylor N Carlson. 2020. "What to believe? Social media commentary and belief in misinformation." *Political Behavior* 42(3):697–718.

Arceneaux, Kevin, Martial Foucault, Kalli Giannelos, Jonathan Ladd and Can Zengin. 2023. The effects of Facebook access during the 2022 French presidential election: Can we incentivize citizens to be better informed and less polarized? Technical report Working Paper.

Arceneaux, Kevin and Martin Johnson. 2013. *Changing minds or changing channels?: Partisan news in an age of choice*. University of Chicago Press.

Aruguete, Natalia, Ernesto Calvo and Tiago Ventura. 2023. "News by popular demand: Ideological congruence, issue salience, and media reputation in news sharing." *The International Journal of Press/Politics* 28(3):558–579.

Arugute, Natalia, Ernesto Calvo and Tiago Ventura. 2022. "Network activated frames: content sharing and perceived polarization in social media." *Journal of Communication* .

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).

Aslett, Kevin, Andrew M Guess, Richard Bonneau, Jonathan Nagler and Joshua A Tucker. 2022. "News credibility labels have limited average effects on news diet quality and fail to reduce misperceptions." *Science advances* 8(18):eabl3844.

Badrinathan, Sumitra and Simon Chauchard. 2023. ""I Don't Think That's True, Bro!" Social Corrections of Misinformation in India." *The International Journal of Press/Politics* 0(0):19401612231158770.

URL: https://doi.org/10.1177/19401612231158770

Badrinathan, Sumitra, Simon Chauchard and Niloufer Siddiqui. 2024. "Misinformation and support for vigilantism: An experiment in India and Pakistan." *American Political Science Review* pp. 1–19.

Banks, Antoine, Ernesto Calvo, David Karol and Shibley Telhami. 2021. "# polarizedfeeds: Three experiments on polarization, framing, and social media." *The International Journal of Press/Politics* 26(3):609–634.

Barberá, Pablo. 2020. "Social Media, Echo Chambers, and Political Polarization." *Social Media and Democracy: The State of the Field, Prospects for Reform* p. 34.

Bessone, Pedro, Filipe R Campante, Claudio Ferraz and Pedro Souza. 2022. Social media and the behavior of politicians: Evidence from Facebook in Brazil. Technical report National Bureau of Economic Research.

Blair, Robert A, Jessica Gottlieb, Brendan Nyhan, Laura Paler, Pablo Argote and Charlene J Stainfield. 2024. "Interventions to counter misinformation: Lessons from the Global North and applications to the Global South." *Current Opinion in Psychology* 55:101732.

Boczkowski, Pablo J, Eugenia Mitchelstein and Mora Matassi. 2018. ""News comes across when I'm in a moment of leisure": Understanding the practices of incidental news consumption on social media." *New media & society* 20(10):3523–3539.

Bor, Alexander and Michael Bang Petersen. 2022. "The psychology of online political hostility: A comprehensive, cross-national test of the mismatch hypothesis." *American political science review* 116(1):1–18.

Bowles, Jeremy, Horacio Larreguy and Shelley Liu. 2020. "Countering misinformation via WhatsApp: Evidence from the COVID-19 pandemic in Zimbabwe." *PloS one* 15(10):e0240005.

Bowles, Jeremy, Kevin Croke, Horacio Larreguy, John Marshall and Shelley Liu. 2025. "Sustaining exposure to fact-checks: Misinformation discernment, media consumption, and its political implications." *American Political Science Review* pp. 1–24.

Budak, Ceren, Brendan Nyhan, David M Rothschild, Emily Thorson and Duncan J Watts. 2024. "Misunderstanding the harms of online misinformation." *Nature* 630(8015):45–53.

Chauchard, Simon and Kiran Garimella. 2022. "What Circulates on Partisan WhatsApp in India? Insights from an Unusual Dataset." *Journal of Quantitative Description: Digital Media* 2.

Chen, Gina Masullo. 2017. *Online incivility and public debate: Nasty talk*. Springer.

Druckman, James N, Erik Peterson and Rune Slothuus. 2013. "How elite partisan polarization affects public opinion formation." *American political science review* 107(1):57–79.

Druckman, James N, Suji Kang, James Chu, Michael N. Stagnaro, Jan G Voelkel, Joseph S Mernyk, Sophia L Pink, Chrystal Redekopp, David G Rand and Robb Willer. 2023. "Correcting misperceptions of out-partisans decreases American legislators' support for undemocratic practices." *Proceedings of the National Academy of Sciences* 120(23):e2301836120.

Ecker, Ullrich KH, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K Fazio, Nadia Brashier, Panayiota Kendeou, Emily K Vraga and Michelle A Amazeen. 2022. "The psychological drivers of misinformation belief and its resistance to correction." *Nature Reviews Psychology* 1(1):13–29.

Flaxman, Seth, Sharad Goel and Justin M Rao. 2016. "Filter bubbles, echo chambers, and online news consumption." *Public opinion quarterly* 80(S1):298–320.

Garimella, Kiran and Dean Eckles. 2020. "Images and misinformation in political groups: Evidence from WhatsApp in India." *arXiv preprint arXiv:2005.09784* .

Ghanem, Dalia, Sarojini Hirshleifer and Karen Ortiz-Becerra. 2023. "Testing attrition bias in field experiments." *Journal of Human resources* .

Gil de Zúñiga, Homero, Alberto Ardèvol-Abreu and Andreu Casero-Ripollés. 2021. "WhatsApp political discussion, conventional participation and activism: exploring direct, indirect and generational effects." *Information, communication & society* 24(2):201–218.

Goyanes, Manuel, Alberto Ardèvol-Abreu and Homero Gil de Zúñiga. 2023. "Antecedents of news avoidance: competing effects of political interest, news overload, trust in news media, and "news finds me" perception." *Digital Journalism* 11(1):1–18.

Graham, Matthew H and Milan W Svolik. 2020. "Democracy in America? Partisanship, polarization, and the robustness of support for democracy in the United States." *American Political Science Review* 114(2):392–409.

Guess, Andrew, Brendan Nyhan and Jason Reifler. 2018. "Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign.".

Guess, Andrew M., Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Edward Kennedy, Young Mie Kim, David Lazer, Devra Moehler, Brendan Nyhan, Carlos Velasco Rivera, Jaime Settle, Daniel Robert Thomas, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud and Joshua A. Tucker. 2023. "How do social media feed algorithms affect attitudes and behavior in an election campaign?" *Science* 381(6656):398–404.
URL: https://www.science.org/doi/abs/10.1126/science.abp9364

Hanley, Sarah M, Susan E Watt and William Coventry. 2019. "Taking a break: The effect of taking a vacation from Facebook and Instagram on subjective well-being." *Plos one* 14(6):e0217743.

Haque, Md Mahfuzul, Mohammad Yousuf, Ahmed Shatil Alam, Pratyasha Saha, Syed Ishtiaque Ahmed and Naeemul Hassan. 2020. "Combating misinformation in Bangladesh: Roles and responsibilities as perceived by journalists, fact-checkers, and users." *Proceedings of the ACM on Human-Computer Interaction* 4(CSCW2):1–32.

Hoes, Emma, Brian Aitken, Jingwen Zhang, Tomasz Gackowski and Magdalena Wojcieszak. 2024.

"Prominent misinformation interventions reduce misperceptions but increase scepticism." *Nature Human Behaviour* 8(8):1545–1553.

Iyengar, Shanto, Gaurav Sood and Yphtach Lelkes. 2012. "Affect, not ideologya social identity perspective on polarization." *Public opinion quarterly* 76(3):405–431.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra and Sean J Westwood. 2019. "The origins and consequences of affective polarization in the United States." *Annual review of political science* 22(1):129–146.

Jenke, Libby. 2024. "Affective polarization and misinformation belief." *Political Behavior* 46(2):825–884.

Jones, Marc Owen. 2022. *Digital authoritarianism in the Middle East: Deception, disinformation and social media*. Oxford University Press.

Kalogeropoulos, Antonis and Patrícia Rossini. 2025. "Unraveling WhatsApp group dynamics to understand the threat of misinformation in messaging apps." *New Media & Society* 27(3):1625–1650.

URL: https://doi.org/10.1177/14614448231199247

Kim, Jin Woo, Andrew Guess, Brendan Nyhan and Jason Reifler. 2021. "The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity." *Journal of Communication* 71(6):922–946.

Kingzette, Jon, James N Druckman, Samara Klar, Yanna Krupnikov, Matthew Levendusky and John Barry Ryan. 2021. "How affective polarization undermines support for democratic norms." *Public Opinion Quarterly* 85(2):663–677.

Kross, Ethan, Philippe Verduyn, Gal Sheppes, Cory K Costello, John Jonides and Oscar Ybarra. 2021. "Social media and well-being: Pitfalls, progress, and next steps." *Trends in cognitive sciences* 25(1):55–66.

Lelkes, Yphtach, Gaurav Sood and Shanto Iyengar. 2017. "The hostile audience: The effect of access to broadband internet on partisan affect." *American Journal of Political Science* 61(1):5–20.

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

Lorenz-Spreen, Philipp, Lisa Oswald, Stephan Lewandowsky and Ralph Hertwig. 2022. "A systematic review of worldwide causal and correlational evidence on digital media and democracy." *Nature human behaviour* pp. 1–28.

Martel, Cameron, Gordon Pennycook and David G Rand. 2020. "Reliance on emotion promotes belief in fake news." *Cognitive research: principles and implications* 5:1–20.

Masip, Pere, Jaume Suau, Carles Ruiz-Caballero, Pablo Capilla and Klaus Zilles. 2021. "News engagement on closed platforms. Human factors and technological affordances influencing exposure to news on WhatsApp." *Digital Journalism* 9(8):1062–1084.

Matthes, Jörg, Kathrin Karsay, Desirée Schmuck and Anja Stevic. 2020. ""Too much to handle": Impact of mobile social networking sites on information overload, depressive symptoms, and well-being." *Computers in Human Behavior* 105:106217.

Newman, Nic, Richard Fletcher, Anne Schulz, Simge Andı, Craig T. Robertson and Rasmus Kleis Nielsen. 2021. "The Reuters Institute Digital News Report 2021." https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2021-06/Digital_News_Report_2021_FINAL.pdf.

Newman, Nic, Richard Fletcher, Craig T Robertson, A Ross Arguedas and Rasmus Kleis Nielsen. 2024. *Reuters Institute digital news report 2024.* Reuters Institute for the study of Journalism.

Nyhan, Brendan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y. Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Andrew M. Guess, Edward Kennedy, Young Mie Kim, David Lazer, Neil Malhotra, Devra Moehler, Jennifer Pan, Daniel Robert Thomas, Rebekah Tromble, Carlos Velasco Rivera, Arjun Wilkins, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud and Joshua A. Tucker. 2023. "Like-minded sources on Facebook are prevalent but not polarizing." *Nature* .

URL: https://doi.org/10.1038/s41586-023-06297-w

Pennycook, Gordon, Tyrone D Cannon and David G Rand. 2018. "Prior exposure increases perceived accuracy of fake news." *Journal of experimental psychology: general* 147(12):1865.

Piazza, James A. 2023. "Political polarization and political violence." *Security Studies* 32(3):476–504.

Poushter, Jacob. 2024. "WhatsApp and Facebook dominate the social media landscape in middle-income nations." *Pew Research Center* .

Rathje, Steve, Jay J Van Bavel and Sander Van Der Linden. 2021. "Out-group animosity drives engagement on social media." *Proceedings of the national academy of sciences* 118(26):e2024292118.

Recuero, Raquel, Felipe Soares and Otávio Vinhas. 2021. "Discursive strategies for disinformation on WhatsApp and Twitter during the 2018 Brazilian presidential election." *First Monday* .

Resende, Gustavo, Philipe Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara Almeida and Fabrício Benevenuto. 2019. (Mis) information dissemination in WhatsApp: Gathering, analyzing and countermeasures. In *The World Wide Web Conference*. pp. 818–828.

Rossini, Patrícia. 2022. "Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk." *Communication Research* 49(3):399–425.

Rossini, Patrícia, Jennifer Stromer-Galley, Erica Anita Baptista and Vanessa Veiga de Oliveira. 2021. "Dysfunctional information sharing on WhatsApp and Facebook: The role of political talk, cross-cutting exposure and social corrections." *New Media & Society* 23(8):2430–2451.

Rowe, Ian. 2015. "Civility 2.0: A comparative analysis of incivility in online political discussion." *Information, communication & society* 18(2):121–138.

Saha, Punyajoy, Binny Mathew, Kiran Garimella and Animesh Mukherjee. 2021. "Short is the Road That Leads from Fear to Hate": Fear Speech in Indian WhatsApp Groups. In *Proceedings of the Web Conference 2021*. p. 1110–1121.

Settle, Jaime E. 2018. *Frenemies: How social media polarizes America*. Cambridge University Press.

Stroud, Natalie Jomini. 2011. *Niche news: The politics of news choice*. Oxford University Press.

Svolik, Milan W. 2019. "Polarization versus democracy." *Journal of democracy* 30(3):20–32.

Tucker, Joshua A, Yannis Theocharis, Margaret E Roberts and Pablo Barberá. 2017. "From liberation to turmoil: Social media and democracy." *J. Democracy* 28:46.

Twenge, Jean M and W Keith Campbell. 2018. "Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study." *Preventive medicine reports* 12:271–283.

Valenzuela, Sebastián, Ingrid Bachmann and Matías Bargsted. 2021. "The personal is the political? What do Whatsapp users share and how it matters for news knowledge, polarization and participation in Chile." *Digital journalism* 9(2):155–175.

Vanman, Eric J, Rosemary Baker and Stephanie J Tobin. 2018. "The burden of online friends: The effects of giving up Facebook on stress and well-being." *The Journal of social psychology* 158(4):496–508.

Velez, Yamil Ricardo and Patrick Liu. 2024. "Confronting core issues: A critical assessment of attitude polarization using tailored experiments." *American Political Science Review* pp. 1–18.

Ventura, Tiago, Rajeshwari Majumdar, Jonathan Nagler and Joshua A. Tucker. 2025. "Misinformation Beyond Traditional Feeds: Evidence from a WhatsApp Deactivation Experiment in Brazil." *The Journal of Politics* .

Voelkel, Jan G, Michael Stagnaro, James Chu, Sophia Pink, Joseph Mernyk, Chrystal Redekopp, Isaias Ghezae, Matthew Cashman, Dhaval Adjodah, Levi Allen et al. 2023. "Megastudy identifying effective interventions to strengthen Americans' democratic attitudes.".

Wilkinson, Steven. 2006. *Votes and violence: Electoral competition and ethnic riots in India*. Cambridge University Press.

Wirtschafter, Valerie, Frederico Batista Pereira, Natália Bueno, Nara Pavão, Felipe Nunes et al. 2024. "Detecting misinformation: Identifying false news spread by political leaders in the global south." *Journal of Quantitative Description: Digital Media* 4.

Statements

Author Contributions

R.M. and T.V. are co-first authors ordered alphabetically. R.M., T.V., S.L., and J.T. designed the study. R.M., T.V., S.L., and C.T. collected the data and supervised the fieldwork. R.M., T.V., S.L., and C.T. wrote the first draft, and all authors contributed to reviewing and editing it.

Funding Statement

This research was funded by the McCourt School's Tech & Policy Program and Project Liberty's Institute. We also gratefully acknowledge that The Center for Social Media and Politics at New York University is supported by funding from the John S. and James L. Knight Foundation, the Charles Koch Foundation, the Hewlett Foundation, Craig Newmark Philanthropies, the William and Flora Hewlett Foundation, the Siegel Family Endowment, and the Bill and Melinda Gates Foundation.

Acknowledgments

We thank Jeremy Bowles, Nejla Asimovic, Anna Wilke, Anjali Thomas, and participants at the 2024 American Political Science Association Annual Meeting, 2025 Midwest Political Science Association Annual Conference, 2025 Media Effects Workshop at Columbia University, 2025 NYU CSMaP Annual Conference, Georgetown Mortara Research Seminar, FGV-EBAPE Seminar Series, and NYU WINNING Summer Workshop for helpful feedback. We also thank Maitreyi Natarajan, Anjali Ofori, Brooks Clifford, Isabella Remor, and Sunaina Kathpalia for excellent research assistance.

Competing Interests

J.T. received a one-time fee from Facebook, the parent company of WhatsApp, to compensate him for administrative time spent in organizing a 1-day conference for approximately 30 academic researchers and a dozen Facebook product managers and data scientists that was held at NYU in the summer of 2017 to discuss research related to civic engagement; the fee was paid before any work or data collection for the current project began. He did not provide any consulting services nor any advice to Facebook as part of this arrangement. J.T. is currently a co-chair of the external academic team for the U.S. 2020 Facebook & Instagram Election Study, a research collaboration between a team of external academic researchers and internal Meta researchers; he receives no financial compensation from Meta for this work. J.T. is currently a Senior Geopolitical Risk Advisor at Kroll. The remaining authors declare no conflicts of interest.

Reducing Social Media Usage During Elections: Evidence from a WhatsApp Multi-Country Experiment

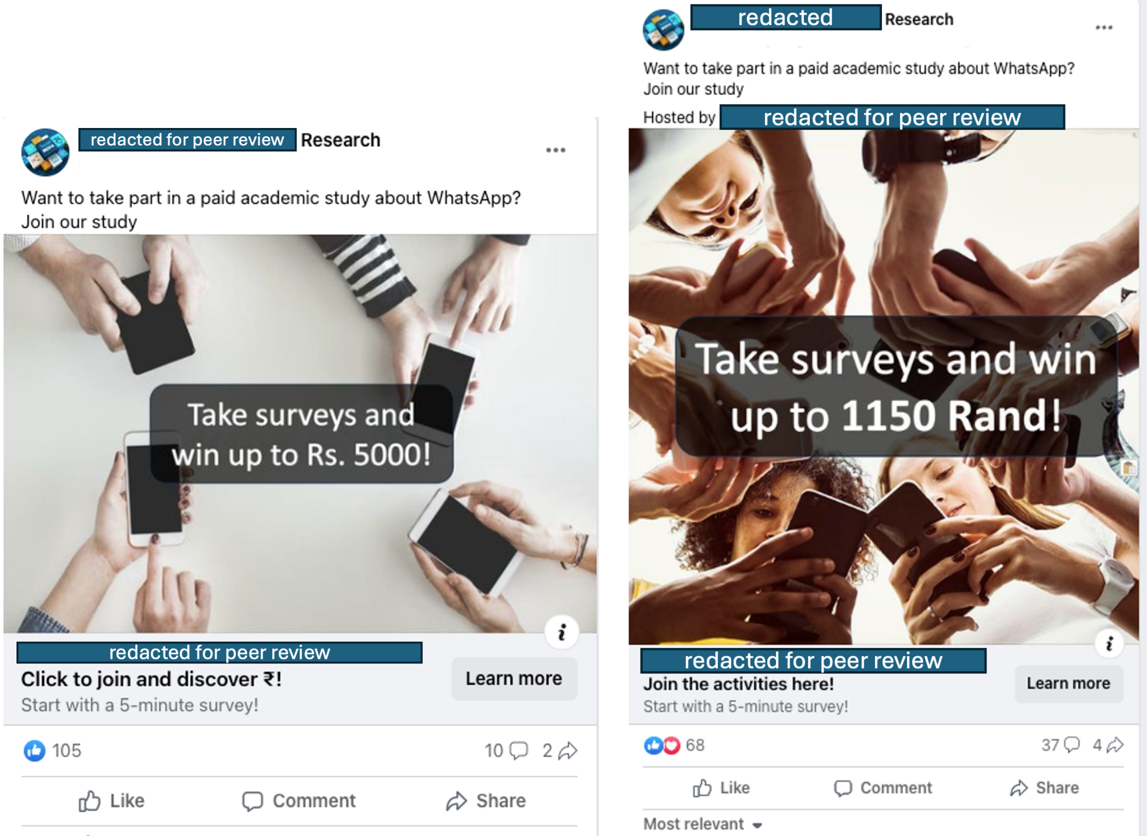
APPENDIX

A Experiment Materials

A.1 Recruitment Materials

For subject recruitment via Meta Advertisements, we used simple text and images to invite users to a paid academic study about WhatsApp. We used multiple small variations of the text and images to increase recruitment. We present examples that respondents saw on their Facebook timelines:

Figure A5: Facebook Advertisement Used for Recruitment



redacted for peer review Research

Want to take part in a paid academic study about WhatsApp?
Join our study

Take surveys and win up to Rs. 5000!

redacted for peer review
Click to join and discover ₹!
Start with a 5-minute survey!

Learn more

105 10 2

Like Comment Share

a) India Ad

redacted Research

Want to take part in a paid academic study about WhatsApp?
Join our study
Hosted by redacted for peer review

Take surveys and win up to 1150 Rand!

redacted for peer review
Join the activities here!
Start with a 5-minute survey!

Learn more

68 37 4

Like Comment Share

Most relevant

b) South Africa Ad

A.2 Screening

Using responses from the recruitment survey, we applied the following criteria to determine eligibility:

- Participants should be 18 years or older;
- Participants should report using WhatsApp for at least 10 minutes per day;
- Participants should not report using WhatsApp on a shared device;
- Participants should have spent at least one minute completing the recruitment survey;
- Participants should not be identified as bots, spammers, or duplicate observations, as determined by Qualtrics security filters and contact information.

A.3 Incentives

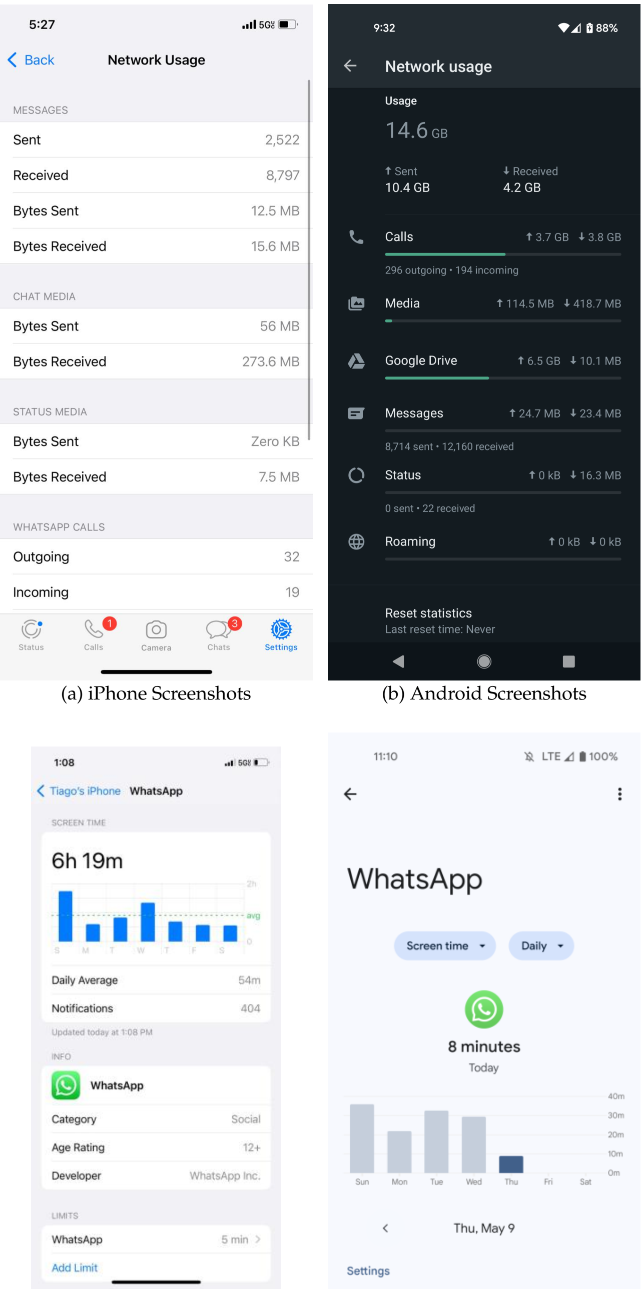
Participants were compensated for every task that they completed. Below, we present the payment values in USD, but participants were paid according to the local currency. Given the differential effort required across treatment and control, we offered an additional bonus for treated participants. We informed participants that this bonus payment is conditional on their compliance with the study, which we would measure by looking at their submitted usage screenshots. The incentives were as follows.

- Baseline survey: 3 dollars
- Compliance task: 1 dollar for each screenshot (4 dollars total)
- Final survey: 8 dollars
- Bonus payment for treated participants who successfully comply: 45 dollars
- Lottery prize: 100 dollars

Thus, those in the control conditions could earn up to 15 dollars; those in the treatment conditions, 60 dollars. All participants were entered into a lottery for one of three 100-dollar gift cards at the end of the study.

A.4 Usage Screenshots

Figure A6: WhatsApp Usage Screenshot



5:27

Back Network Usage

MESSAGES

Sent	2,522
Received	8,797
Bytes Sent	12.5 MB
Bytes Received	15.6 MB

CHAT MEDIA

Bytes Sent	56 MB
Bytes Received	273.6 MB

STATUS MEDIA

Bytes Sent	Zero KB
Bytes Received	7.5 MB

WHATSAPP CALLS

Outgoing	32
Incoming	19

(a) iPhone Screenshots

9:32 88%

Network usage

Usage

14.6 GB

↑ Sent 10.4 GB

↓ Received 4.2 GB

Calls	↑ 3.7 GB	↓ 3.8 GB
Media	↑ 114.5 MB	↓ 418.7 MB
Google Drive	↑ 6.5 GB	↓ 10.1 MB
Messages	↑ 24.7 MB	↓ 23.4 MB
Status	↑ 0 kB	↓ 16.3 MB
Roaming	↑ 0 kB	↓ 0 kB

296 outgoing • 194 incoming

8,714 sent • 12,160 received

0 sent • 22 received

Reset statistics

Last reset time: Never

(b) Android Screenshots

1:08

Tiago's iPhone WhatsApp

SCREEN TIME

6h 19m

Daily Average	54m
Notifications	404

Updated today at 1:08 PM

INFO

WhatsApp

Category	Social
Age Rating	12+
Developer	WhatsApp Inc.

LIMITS

WhatsApp	5 min

Add Limit

(c) iPhone Screenshots

11:10 LTE 100%

WhatsApp

Screen time ▾ Daily ▾

8 minutes

Today

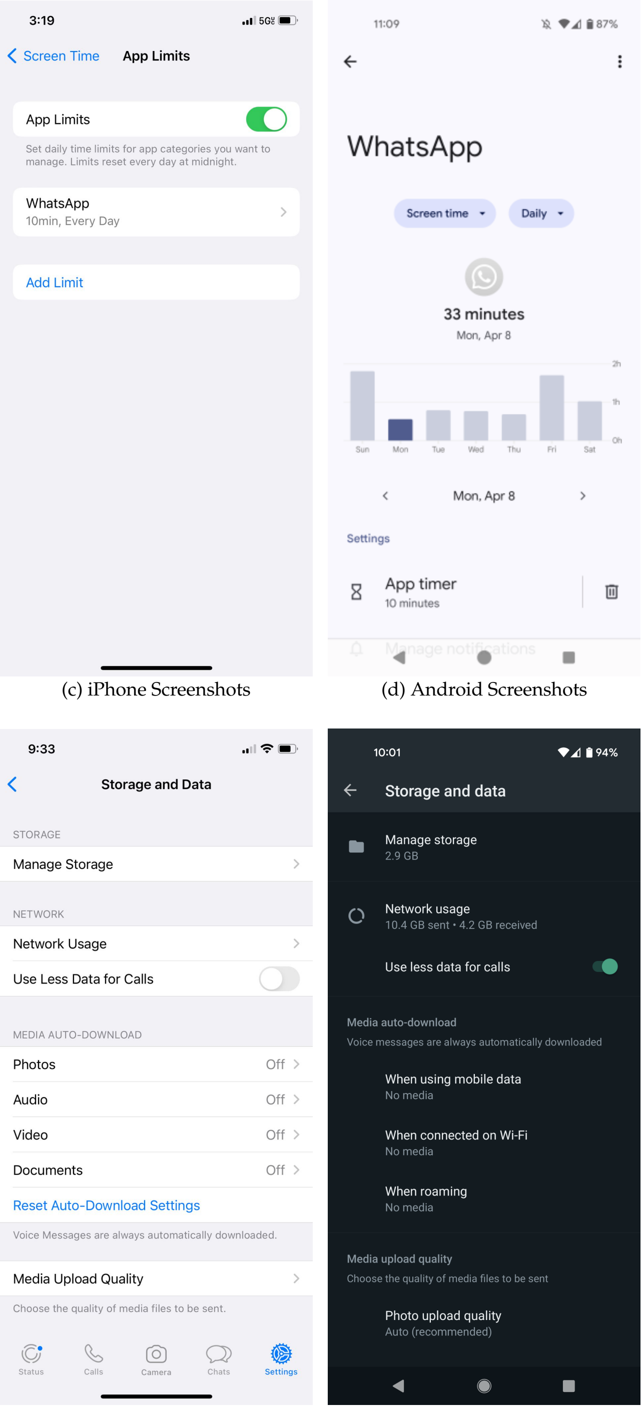
Thu, May 9

Settings

(d) Android Screenshots

Note: The upper row presents examples of the screenshots of multimedia consumption requested from participants assigned to **Multimedia**. The bottom row presents examples of the screenshots of usage information requested from participants assigned to **Time**.

Figure A7: WhatsApp Treatment Screenshot



3:19

Screen Time App Limits

App Limits ●

Set daily time limits for app categories you want to manage. Limits reset every day at midnight.

WhatsApp
10min, Every Day

Add Limit

(c) iPhone Screenshots

11:09 87%

WhatsApp

Screen time ▾ Daily ▾

33 minutes
Mon, Apr 8

Mon, Apr 8

Settings

App timer
10 minutes

Manage notifications

(d) Android Screenshots

9:33

Storage and Data

STORAGE

Manage Storage

NETWORK

Network Usage

Use Less Data for Calls ○

MEDIA AUTO-DOWNLOAD

Photos	Off
Audio	Off
Video	Off
Documents	Off

Reset Auto-Download Settings

Voice Messages are always automatically downloaded.

Media Upload Quality

Choose the quality of media files to be sent.

(a) iPhone Screenshots

10:01 94%

Storage and data

Manage storage
2.9 GB

Network usage
10.4 GB sent • 4.2 GB received

Use less data for calls ●

Media auto-download
Voice messages are always automatically downloaded

When using mobile data
No media

When connected on Wi-Fi
No media

When roaming
No media

Media upload quality
Choose the quality of media files to be sent

Photo upload quality
Auto (recommended)

(b) Android Screenshots

Note: The upper row presents examples of the screenshots requested from participants assigned to **Time**. The bottom row presents examples of the screenshots requested from participants assigned to **Multimedia**.

B Outcome Measures

Tables B2 and B3 describe the outcomes included in our primary analyses. Table B4 presents the news and misinformation headlines shown to participants in each country. All headlines are presented in English as seen by participants in India and South Africa. Participants in Brazil saw them in Portuguese, but we present their translated versions here.

Table B2: Pre-Registered Outcomes and Measurement Choices for Information Outcomes

Variable Name	**Definition**	**Measurement**
Misinformation Recall	Measures the self-reported recall of Misinformation Rumors using a headline task	Sum of the Misinformation Rumors items recalled.
News Recall	Measures the self-reported recall of True News using a headline task	Sum of the True News items recalled.
Misinformation Accuracy	Measures the ability to discern false information using a headline task	Sum of the Misinformation Rumors classified as false.
News Knowledge Accuracy	Measures the ability to discern true news using a headline task	Sum of the news headlines, True News and Placebo False News,correctly classified as accurate or not accurate, respectively.
Online Toxicity	Composite social media incivility score	Sum of the standardized *z*-scores for online toxicity and online incivility measures
Quality of Political Discussions	Composite score for experiences of the quality of political discussion	Sum of the standardized *z*-scores for four items: political anger, political incivility, information overload

Table B3: Pre-Registered Outcomes and Measurement Choices for Political Outcomes

Variable Name	**Definition**	**Measurement**
Polarization Index	Composite polarization measure	Sum of the z-scores for three polarization outcomes (affective polarization, social polarization, traits polarization)
Affective Polarization	Measures affective polarization towards the two major political parties/coalition	Absolute value of the difference between the feeling scales for each political party/coalition
Social Polarization	Measures social polarization towards the two major political parties/coalition	Absolute value of the difference between the number of social activities willing to engage with voters of each political party/coalition
Traits Polarization	Measures positive and negative traits assigned to voters of main political parties/coalition	Difference between number of negative and positive traits assigned to outgroup party/coalition
Identity-based prejudice Index	Composite index measuring levels of identity-based prejudice	Sum of the z-scores for three identity polarization measures index (Affective prejudice, Social prejudice, Traits prejudice)
Affective Identity-Based Prejudice	Measures affective outgroup prejudice between participants' ingroup and outgroup in each country	Absolute value of the difference between the feeling scales for in/outgroup
Social Identity-Based Prejudice	Measures social prejudice between participants' ingroup and outgroup in each country	Absolute value of the difference between the number of social activities willing to engage with ingroup and outgroup
Traits Identity-Based Prejudice	Measures positive and negative traits assigned to participants' outgroup in each country	Difference between number of negative and positive traits assigned to outgroup
Candidate Favorability	Measure overall positive feeling towards the participants' preferred candidate	Absolute difference between participants' main political ingroup candidate and average favorability of the candidates from the alternative parties
Issue Polarization	Measures extremity compared to average voter with respect to six issue opinions questions	Standardized absolute differences between participants' agreement (on each issue) to the mean issue agreement among the entire control group

Table B4: Misinformation Rumors and True News Headlines

Headlines	Category
Brazil	
At a rally on September 7, Bolsonaro got angry at Pablo Marçal's sound car and called the candidate a lazy person	Misinformation Rumors
In a video, journalist Sandra Annenberg announced the resgata brasil program, where Brazilians can win up to 7,000 reais in compensation due to a leakage of credit card data	Misinformation Rumors
In a recent decision, Finance Minister Fernando Haddad announced 0nd of unemployment insurance for 2025	Misinformation Rumors
Commenting on the elections in Venezuela, Lula criticizes the electronic ballot box, and defends the use of printed ballots by Venezuelan Authorities	Misinformation Rumors
For drug trafficking and corruption with public money, court orders arrest of singer Gustavo Lima and influencer Deolane Bezerra	Placebo News
After negotiations in New York, Israel, Lebanon and Palestine reach peace agreement and cease attacks in the region.	Placebo News
Government announces that 600 betting sites will be banned from Brazil in October	True News
In New York, President Lula receives award from billionaire Bill Gates for policies to combat hunger	True News
Alexandre de Moraes last week denied Twitter/X's return to Brazil despite the company complying with some of the court's requests	True News
Filhes desse solo: National Anthem is sung in neutral language at Guilherme Boulos' rally with Lula.	True News
India	
Muslim women were caught doing fake voting under their burqas during the Lok Sabha elections	Misinformation Rumors
Congress promises that the money of regular Indian citizens will be collected and distributed to poor Muslims	Misinformation Rumors
No new Public Sector Enterprises have been incorporated under the Modi Govt	Misinformation Rumors
BJP is using Army personnel to influence citizens in election booths to vote for the BJP	Misinformation Rumors
Rahul Gandhi praises PM Modi for his quick response to Pune Porsche accident	Placebo News
New study finds population rise is related to religion, highest increase in fertility rate among Muslims	Placebo News
Modi accuses Opposition of using "vote jihad" to win elections	True News
Arvind Kejriwal will have to surrender and go back to jail on June 2	True News
Congress leaders call PM Modi a dictator	True News
Election-time seizures of cash, drugs, liquor to cross all-time high mark in 2024	True News
South Africa	
If you are registered the vote but don't vote on elections day, then the ANC will receive your vote automatically.	Misinformation Rumors
Mozambican migrants are being imported into South Africa to vote for the ANC.	Misinformation Rumors
You can get a South African ID for R4,500 quickly by applying through WhatsApp.	Misinformation Rumors
South Africa's biggest trade union, NUMSA, has asked its members to vote for the MK Party.	Misinformation Rumors
Political leaders have called DA's burning of SA flag a 'heroic act'.	Placebo News
Jacob Zuma agreed to step down from the race after being locked out of South Africa elections.	Placebo News
Parliament gives Ramaphosa a blank cheque to set donation limits	True News
DA accuses Rise Mzansi of fueling racial tensions in Western Cape.	True News
The South Africa Medical Association is planning to mount legal challenge against the new National Health Insurance law.	True News
Jabulani Khumalo takes fight to remove Jacob Zuma as MK Party leader to Electoral Court.	True News

C Sample Characteristics and Attrition Analysis

Following the recruitment procedures described in Section 3.1, we invited 6,261 individuals to participate in our WhatsApp usage reduction experiment (2,067 in BR, 1,310 in IN, 2,884 in SA). Out of the individuals invited, 2,425 participants were successfully enrolled in the experiment (926 in Brazil, 679 in India, and 820 in South Africa). Then, out of the participants who started the interventions, 2,220 completed the post-treatment survey (825 in Brazil, 653 in India, and 742 in South Africa). We divide this section into three parts. First, we present the baseline characteristics of our enrolled sample with benchmarks from other population surveys in Brazil, India, and South Africa. Second, we compare baseline characteristics between the participants we invited and those who enrolled in the experiment. We called this analysis *selection among the enrolled*. Understanding this pattern helps us calibrate the external validity of our findings. Lastly, we conduct a series of attrition analyses comparing baseline differences between participants who successfully enrolled in the study versusparticipants who attrited from the study between enrollment and the post-treatment survey. We present a set of formal tests of attrition bias as suggested in Ghanem, Hirshleifer and Ortiz-Becerra (2023).

C.1 Baseline Characteristics

Table C5: Demographics and Political Attitudes, Sample Versus National Benchmarks (Brazil)

	This Study	BES	Online Survey
Age			
18-24	0.17	0.15	0.12
25-34	0.41	0.22	0.31
35-44	0.24	0.21	0.27
45-54	0.13	0.18	0.17
55-64	0.05	0.15	0.08
65+	0.01	0.08	0.04
Gender			
Female	0.65	0.51	0.51
Male	0.35	0.49	0.49
Education			
Less than High School	0.03	0.41	0.08
High School	0.50	0.36	0.27
Professional Degree	0.39	0.09	0.15
College Degree or More	0.07	0.14	0.50
Region			
Centro-Oeste	0.05	0.08	0.07
Nordeste	0.15	0.27	0.23
Norte	0.04	0.08	0.03
Sudeste	0.65	0.43	0.51
Sul	0.11	0.14	0.16
Vote First Round			
Bolsonaro	0.32	0.34	N/A
Lula	0.39	0.38	N/A
Others	0.29	0.29	N/A

Note: Column 1 lists categories from demographics and political variables. Column 2 reports their proportions for participants who successfully enrolled in our study. Column 3 reports benchmark values based on the Brazilian Election Survey from 2022, which consists of an in-person, face-to-face household survey, representative of the Brazilian population. Column 4 reports demographics from a online-based panel of WhatsApp users.

Table C6: Demographics and Political Attitudes, Sample Versus National Benchmarks (India)

	This Study	CVoter	NES
Age			
18-24	0.27	0.16	0.13
25-34	0.4	0.33	0.26
35-44	0.21	0.25	0.23
45-54	0.08	0.16	0.18
55+	0.03	0.1	0.21
Gender			
Female	0.32	0.48	0.49
Male	0.68	0.52	0.51
Education			
Less than High School	0.01	0.45	0.64
High School	0.16	0.19	0.22
College/Professional Degree	0.82	0.36	0.14
Region			
East	0.14	0.26	N/A
North/Central	0.39	0.38	N/A
Northeast	0.02	0.03	N/A
South	0.27	0.23	N/A
West	0.18	0.14	N/A
Religion			
Buddhism	0.02	N/A	N/A
Christianity	0.04	N/A	0.02
Hinduism	0.79	N/A	0.8
Islam	0.13	N/A	0.14
Jainism	0.01	N/A	N/A
Sikhism	0.01	N/A	0.02
Other/Unknown	N/A	N/A	0.01
Preferred Coalition			
BJP Coalition	0.69	N/A	0.55
Congress Coalition	0.31	N/A	0.45
Level of BJP Support			
Somewhat/Strong	0.45	0.48	0.36
Other	0.54	0.52	0.64

Note: Column 1 lists categories from demographics and political variables. Column 2 reports their proportions for participants who successfully enrolled in our study. Column 3 reports benchmark values based on a 2024 survey conducted via CVoter, which consists of telephone survey representative of the Indian population. Column 4 reports benchmark values based on the 2024 Indian National Election Study conducted by Lokniti-CSDS.

Table C7: Demographics and Political Attitudes, Sample Versus National Benchmarks (South Africa)

	This Study	Afrobarometer	
Age			
18-24	0.16	0.18	0.22
25-34	0.51	0.23	0.29
35-44	0.23	0.22	0.25
45-54	0.07	0.17	0.15
55-64	0.01	0.12	0.07
65+	0.01	0.08	0.03
Gender			
Female	0.71	0.50	0.49
Male	0.29	0.50	0.51
Education			
College/Professional Degree	0.57	0.00	0.28
High School	0.39	0.00	0.44
Less than High School	0.05	0.00	0.28
Region			
Eastern Cape	0.05	0.10	0.09
Free State	0.04	0.07	0.07
Gauteng	0.46	0.25	0.27
KwaZulu-Natal	0.15	0.16	0.16
Limpopo	0.08	0.09	0.08
Mpumalanga	0.05	0.07	0.07
North West	0.04	0.07	0.70
Northern Cape	0.01	0.06	0.05
Western Cape	0.13	0.12	0.13
Would Vote For			
African National Congress	0.31	0.32	0.29
Democratic Alliance	0.18	0.11	0.12
Others	0.51	0.17	0.19

Note: Column 1 lists categories from demographics and political variables. Column 2 reports their proportions for participants who successfully enrolled in our study. Column 3 reports values based on the Afrobarometer Survey Round 9 (2022). Column 4 reports values based on the Afrobarometer Survey, conditioning on respondents who indicated that they use social media for news.

C.2 Selection Among the Enrolled

In this section, we report differences in the sample characteristics between individuals invited to join the experiment and those who successfully enrolled. Invited participants are those who passed all of the eligibility criteria listed in Appendix Section A.2, and we sent at least one email or WhatsApp message inviting them to return to the study. All eligible participants in India and

Table C8: Baseline Characteristics Between Invited and Enrolled Participants

	Dropout (N=3836)		Enrolled (N=2425)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
Age	2.62	1.22	2.35	1.05	-0.27	<0.01
Gender	0.42	0.50	0.43	0.50	0.01	0.60
Education	4.47	0.92	4.69	0.85	0.22	<0.01
WP: Daily time	3.49	1.48	3.61	1.40	0.13	<0.01
WP: Mobile Only	0.86	0.35	0.71	0.46	-0.15	<0.01
WP: Use for Work	5.12	1.17	5.09	1.05	-0.02	0.40
WP: Use for News	4.99	1.23	4.78	1.27	-0.21	<0.01
WP: Use for Chatting with Friends	5.51	0.88	5.44	0.87	-0.06	<0.01
WP: Frequency Images about Politics	3.92	1.54	4.14	1.47	0.21	<0.01
Omnibus Test: F-Statistics = 49.8 *p*-values <0.01						

South Africa were invited, while in Brazil, we achieved our target sample size before inviting all eligible individuals. We use the following variables (all converted to numerical scales) to understand differences between these groups: age, gender, education, time spent on WhatsApp, using WhatsApp on mobile only, WhatsApp usage for chatting with friends, WhatsApp usage for news consumption, WhatsApp usage for work-related tasks, and exposure to multimedia content about politics on WhatsApp.

Table C8 presents the results. For each of the eight variables, we provide pairwise t-tests and omnibus F-statistics for the entire model to examine the joint orthogonality for all eight variables. We detect statistically significant differences for most of the variables. Overall, participants who enrolled in the experiments were younger, more educated, used to spend more time on WhatsApp, were less likely to be mobile-only WhatsApp users, less likely to use WhatsApp for news and to talk with friends, but more likely to receive political content through multimedia on WhatsApp. These differences contribute to high and statistically significant F-statistics.

We use the same set of variables to examine the differences between the treatment and control groups of enrolled participants. Table C9 shows these results, and we do not find statistically significant differences in any of the eight variables, or in the omnibus F-test. Therefore, although we observe evidence of selection into the experiment, particularly among a likely more digitally savvy group, we do not find any indication that this issue affects the balance between treatment and control. When considering the external validity of our experiments, these differences indicate our inferences are valid to a greater degree for a sample of heavy WhatsApp users

Table C9: Baseline Characteristics Between Treatment and Control Participants Enrolled

	Control (N=1198)		Treatment (N=1227)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
Age	2.36	1.08	2.34	1.03	-0.02	0.72
Gender	0.42	0.50	0.44	0.51	0.02	0.32
Education	4.71	0.85	4.67	0.84	-0.04	0.23
WP: Daily time	3.63	1.40	3.60	1.41	-0.02	0.67
WP: Mobile Only	0.70	0.46	0.71	0.45	0.02	0.36
WP: Use for Work	5.11	1.04	5.07	1.06	-0.04	0.35
WP: Use for News	4.79	1.27	4.77	1.28	-0.01	0.78
WP: Use for Chatting with Friends	5.44	0.88	5.45	0.85	0.01	0.89
WP: Frequency Images about Politics	4.14	1.49	4.13	1.46	-0.01	0.90
Omnibus Test: F-Statistics = 0.49, *p*-values = 0.893						

C.3 Attrition Analysis

In this section, we report findings related to attrition rates within the intervention. To address the potential for attrition bias, we apply statistical tests introduced by Ghanem, Hirshleifer and Ortiz-Becerra (2023). Rather than focusing solely on baseline comparisons between the treatment and control groups, this approach offers a more formal framework for evaluating attrition bias in field experiments by leveraging baseline outcome data. Specifically, Ghanem, Hirshleifer and Ortiz-Becerra (2023) demonstrate that the key identifying assumption for estimating local treatment effects can be assessed by testing two equality conditions: one concerning the baseline outcome distributions of the treatment and control groups, and another concerning the same distributions among those who attrited from each group. We use the following baseline covariates and proxies for the outcomes: age, gender, education, income, self-reported WhatsApp usage per day, news consumption, news consumption on social media, self-reported exposure to false information, partisan affective polarization, and ethnic prejudice. The last two are calculated using the absolute difference between the two main political and ethnic/racial groups across the three countries, except for Brazil where we did not measure the ethnic prejudice at baseline.

Table C10 examines the presence of attrition bias using this approach. Column 1 reports the attrition rate for control, and Column 2 reports the differential attrition rate between treatment and control, with the corresponding p-value testing for the difference in attrition between the groups (*differential attrition*). We find no evidence of differential attrition in the full sample. Columns 3-6 present the mean baseline outcome for treatment respondents (TR), control respondents (CR), treatment attriters (TA), and control attriters (CA), respectively. Column 7 reports the p-value of the hypothesis test with two equality restrictions (*selective attrition*). We cannot reject the joint null hypothesis of no differences in the mean baseline values of outcome variables across treatment and control respondents, or across treatment and control attritors. This test indicates that selective attrition is not a threat to the experiment's internal validity.

Tables C11, C12, C13 present the same results by country. In Brazil, only one out of the ten variables do not pass the selective attrition test. In India, we find no evidence of differential attrition or selective attrition. In South Africa, no evidence of selective attrition is detected, but we find small variation vis-à-vis differential attrition. Since most of the pooled results are not driven by the South African sample, we do not consider this a threat to the internal validity of our design.

Table C10: Full Sample Internal Validity Test for Primary Outcomes and Baseline Variables

	Attrition Bias		Mean baseline outcome by group				Test of internal validity
Outcome	C	Differential	TR	CR	TA	CA	p-value
Age	0.91	0.002 (p-value = 0.8577)	2.33	2.35	2.43	2.44	0.94
Gender	0.91	0.002 (p-value = 0.8577)	0.44	0.42	0.47	0.44	0.60
Education	0.91	0.002 (p-value = 0.8577)	4.68	4.72	4.51	4.58	0.47
Income	0.91	0.002 (p-value = 0.8577)	4.54	4.64	3.88	3.72	0.36
WP: Daily time	0.91	0.002 (p-value = 0.8577)	3.61	3.65	3.49	3.37	0.68
News Consumption: General	0.91	0.002 (p-value = 0.8577)	3.57	3.61	2.81	2.82	0.77
News Consumption: Social Media Apps	0.91	0.002 (p-value = 0.8577)	0.93	0.92	0.90	0.87	0.57
False News Exposure	0.91	0.002 (p-value = 0.8577)	2.49	2.44	2.51	2.50	0.57
Affective Polarization	0.91	0.002 (p-value = 0.8577)	-0.03	0.01	0.23	0.01	0.16
Ethnic Prejudice	0.91	0.002 (p-value = 0.8577)	0.11	0.16	0.07	0.18	0.61

Note: Column 1 reports the attrition rate for control, and Column 2 reports the differential attrition rate between treatment and control, with the corresponding p-value testing for difference in attrition between the groups (*differential attrition*). Columns 3-6 present the mean baseline outcome for treatment respondents (TR), control respondents (CR), treatment attriters (TA), and control attriters (CA), respectively. Column 7 reports the p-value of the hypothesis test with two equality restrictions (*selective attrition*).

Table C11: Brazil Sample Internal Validity Test for Primary Outcomes and Baseline Variables

	Attrition Bias		Mean baseline outcome by group				Test of internal validity
Outcome	C	Differential	TR	CR	TA	CA	p-value
Age	0.87	0.0317 (p-value = 0.1236)	2.50	2.51	2.63	2.56	0.95
Gender	0.87	0.0317 (p-value = 0.1236)	0.35	0.35	0.44	0.41	0.94
Education	0.87	0.0317 (p-value = 0.1236)	4.49	4.52	4.30	4.46	0.40
Income	0.87	0.0317 (p-value = 0.1236)	5.00	5.13	4.30	3.98	0.38
WP: Daily time	0.87	0.0317 (p-value = 0.1236)	3.79	3.82	3.51	3.41	0.89
News Consumption: General	0.87	0.0317 (p-value = 0.1236)	3.11	3.12	2.51	2.69	0.78
News Consumption: Social Media Apps	0.87	0.0317 (p-value = 0.1236)	0.91	0.87	0.88	0.81	0.11
False News Exposure	0.87	0.0317 (p-value = 0.1236)	2.42	2.36	2.21	2.39	0.60
Affective Polarization	0.87	0.0317 (p-value = 0.1236)	0.19	0.22	0.71	0.14	0.03

Note: Column 1 reports the attrition rate for control, and Column 2 reports the differential attrition rate between treatment and control, with the corresponding p-value testing for difference in attrition between the groups (*differential attrition*). Columns 3-6 present the mean baseline outcome for treatment respondents (TR), control respondents (CR), treatment attritters (TA), and control attritters (CA), respectively. Column 7 reports the p-value of the hypothesis test with two equality restrictions (*selective attrition*).

Table C12: India Sample Internal Validity Test for Primary Outcomes and Baseline Variables

	Attrition Bias		Mean baseline outcome by group				Test of internal validity
Outcome	C	Differential	TR	CR	TA	CA	p-value
Age	0.95	0.0187 (p-value = 0.205)	2.22	2.18	2.10	2.56	0.51
Gender	0.95	0.0187 (p-value = 0.205)	0.69	0.67	0.60	0.50	0.75
Education	0.95	0.0187 (p-value = 0.205)	4.98	4.97	4.30	4.94	0.28
Income	0.95	0.0187 (p-value = 0.205)	4.58	4.49	4.80	4.12	0.31
WP: Daily time	0.95	0.0187 (p-value = 0.205)	3.11	3.15	3.10	2.81	0.81
News Consumption: General	0.95	0.0187 (p-value = 0.205)	4.07	3.99	3.20	2.81	0.57
News Consumption: Social Media Apps	0.95	0.0187 (p-value = 0.205)	0.94	0.96	0.90	0.94	0.39
False News Exposure	0.95	0.0187 (p-value = 0.205)	2.54	2.38	3.00	2.56	0.15
Affective Polarization	0.95	0.0187 (p-value = 0.205)	-0.10	0.05	-0.17	-0.07	0.13

Note: Column 1 reports the attrition rate for control, and Column 2 reports the differential attrition rate between treatment and control, with the corresponding p-value testing for difference in attrition between the groups (*differential attrition*). Columns 3-6 present the mean baseline outcome for treatment respondents (TR), control respondents (CR), treatment attriters (TA), and control attriters (CA), respectively. Column 7 reports the p-value of the hypothesis test with two equality restrictions (*selective attrition*).

Table C13: South Africa Sample Internal Validity Test for Primary Outcomes and Baseline Variables

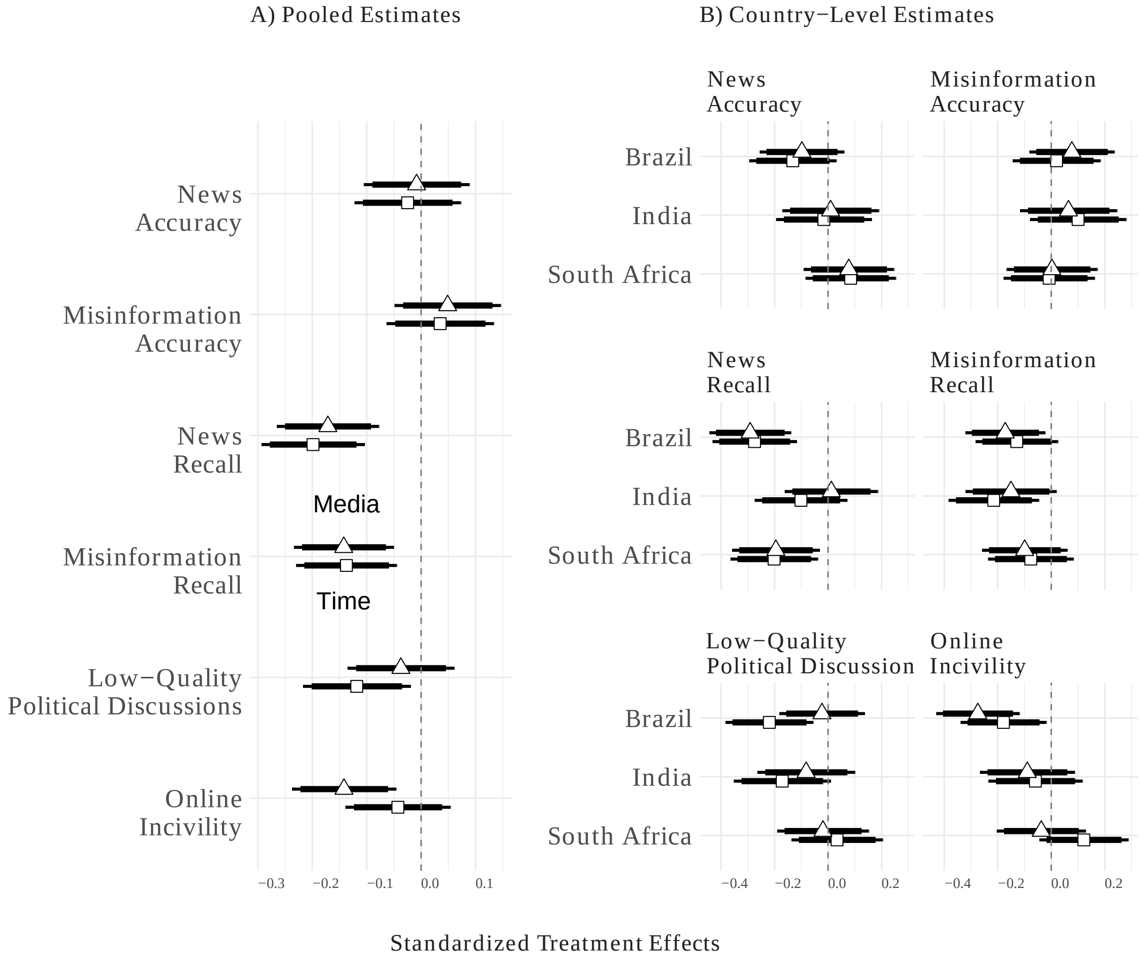
	Attrition Bias		Mean baseline outcome by group				Test of internal validity
Outcome	C	Differential	TR	CR	TA	CA	p-value
Age	0.93	-0.0463 (p-value = 0.0239)	2.25	2.32	2.32	2.11	0.42
Gender	0.93	-0.0463 (p-value = 0.0239)	0.30	0.27	0.46	0.46	0.63
Education	0.93	-0.0463 (p-value = 0.0239)	4.64	4.73	4.74	4.64	0.25
Income	0.93	-0.0463 (p-value = 0.0239)	4.00	4.22	3.34	2.93	0.26
WP: Daily time	0.93	-0.0463 (p-value = 0.0239)	3.85	3.89	3.54	3.61	0.92
News Consumption: General	0.93	-0.0463 (p-value = 0.0239)	3.62	3.83	2.98	3.07	0.14
News Consumption: Social Media Apps	0.93	-0.0463 (p-value = 0.0239)	0.94	0.94	0.92	0.96	0.73
False News Exposure	0.93	-0.0463 (p-value = 0.0239)	2.54	2.58	2.68	2.68	0.90
Affective Polarization	0.93	-0.0463 (p-value = 0.0239)	-0.23	-0.26	-0.10	-0.21	0.73

Note: Column 1 reports the attrition rate for control, and Column 2 reports the differential attrition rate between treatment and control, with the corresponding p-value testing for difference in attrition between the groups (*differential attrition*). Columns 3-6 present the mean baseline outcome for treatment respondents (TR), control respondents (CR), treatment attriters (TA), and control attriters (CA), respectively. Column 7 reports the p-value of the hypothesis test with two equality restrictions (*selective attrition*).

D Additional Results

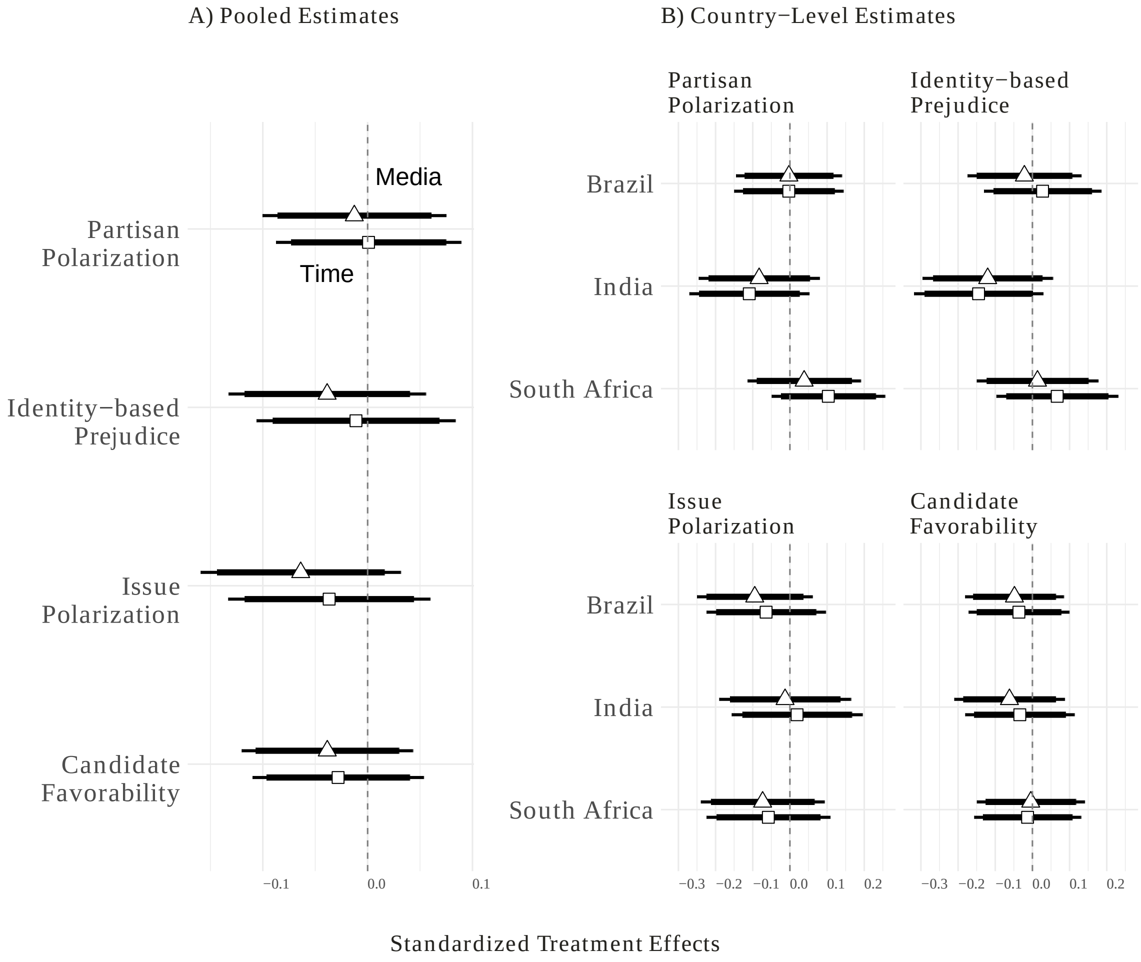
D.1 Unpooled *Multimedia* and *Time* Treatment Effects

Figure D8: Unpooled Treatment Effects on Information Outcomes



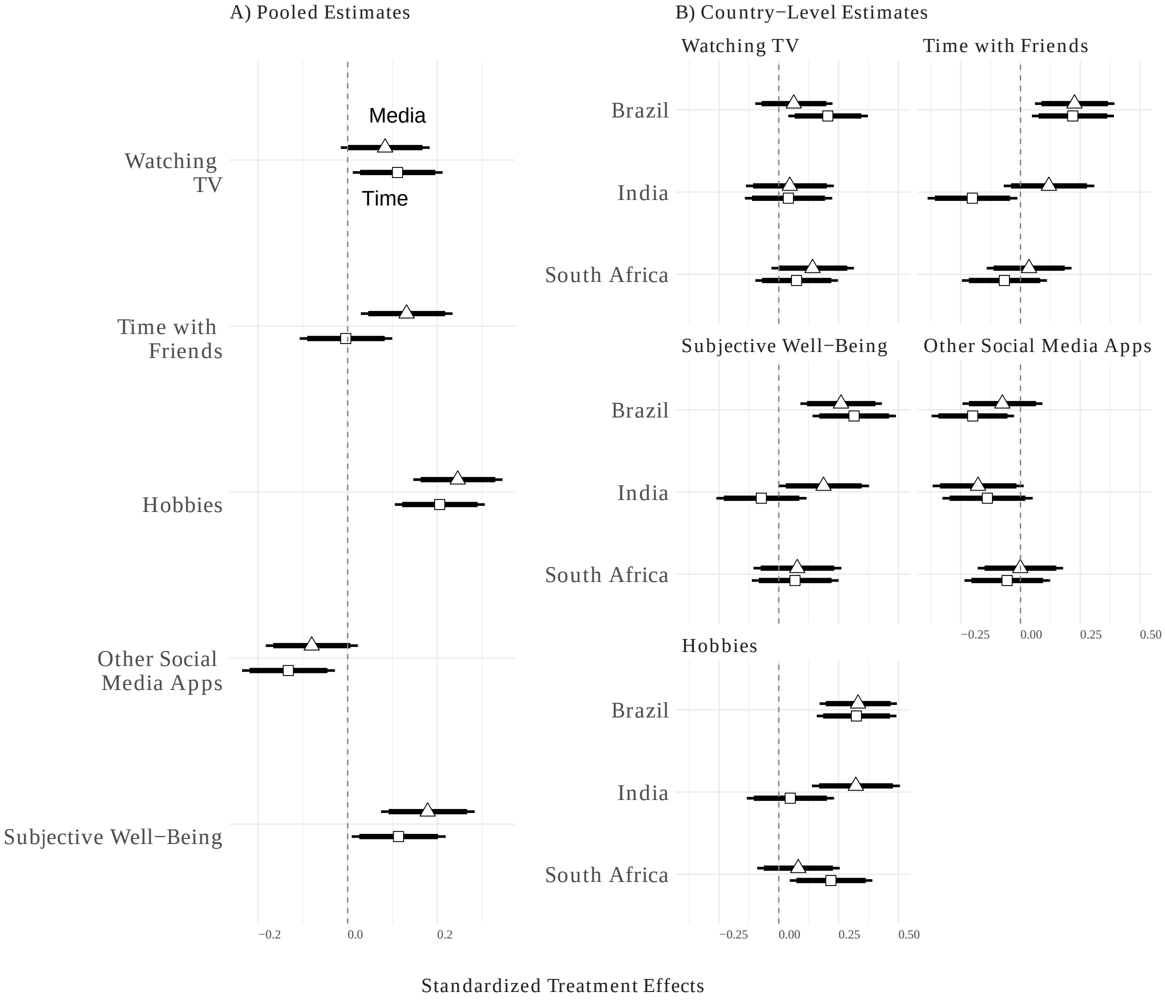
Notes: See Appendix Tables G25 and G26 for full regression results.

Figure D9: Unpooled Treatment Effects on Political Outcomes



Notes: See Appendix Tables G25 and G26 for full regression results.

Figure D10: Unpooled Treatment Effects on Substitutes and Well-Being

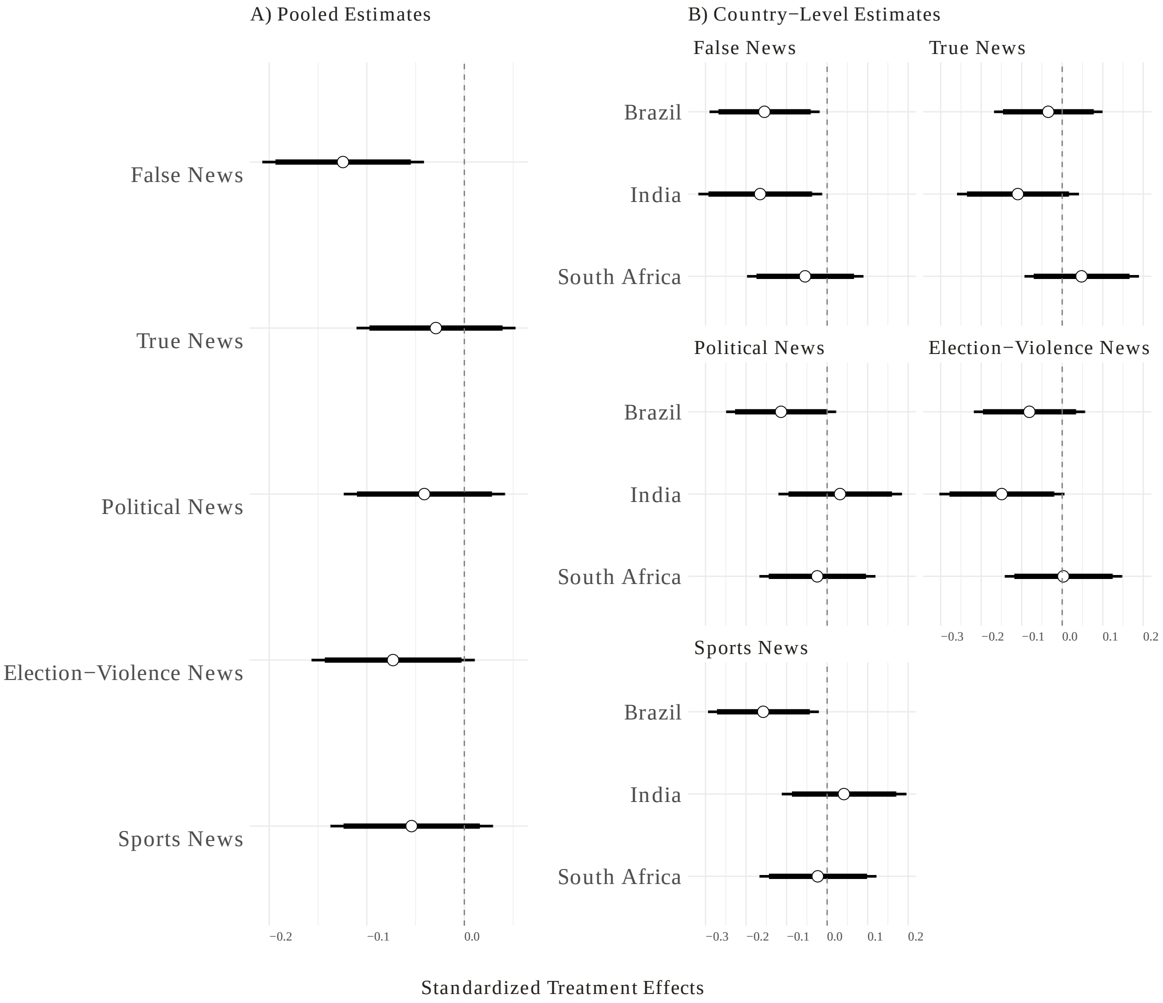


Notes: See Appendix Tables G29 and G30 for full regression results.

D.2 Treatment Effects on Self-Reported Exposure to News

In addition to measuring changes in information consumption using the recall of specific true news and false rumors headlines, we also asked participants directly about the type of information they consumed during the weeks of the intervention. Specifically, we asked: ”Thinking back over the past one month, how much did you see the following on social media?” for the items: (a) information you think is true; (b) information you think is false; (c) news about politics; (d) news about election related violence; and (e) news about sports. Responses were collected on a five-point scale ranging from “never” to “very frequently.” These outcomes help us understand how participants perceive changes in their informational environment after reducing social media usage; they do not replace our pre-registered results using the headlines recall task. Figure D11 presents the results.

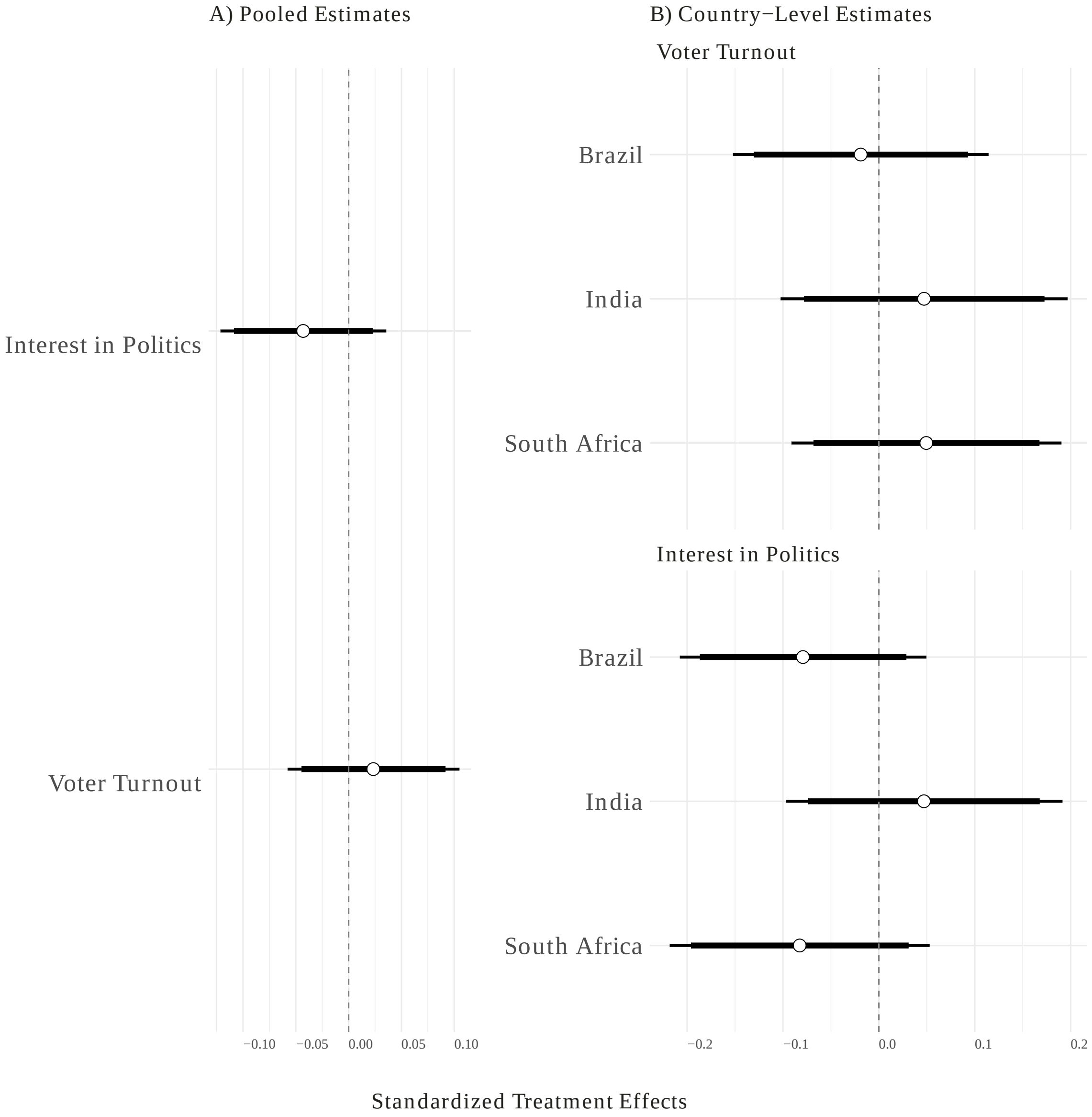
Figure D11: Treatment Effects on Self-Reported Exposure to News Content



Notes: See Appendix Tables G33 and G34 for full regression results.

D.3 Political Interest and Voter Turnout

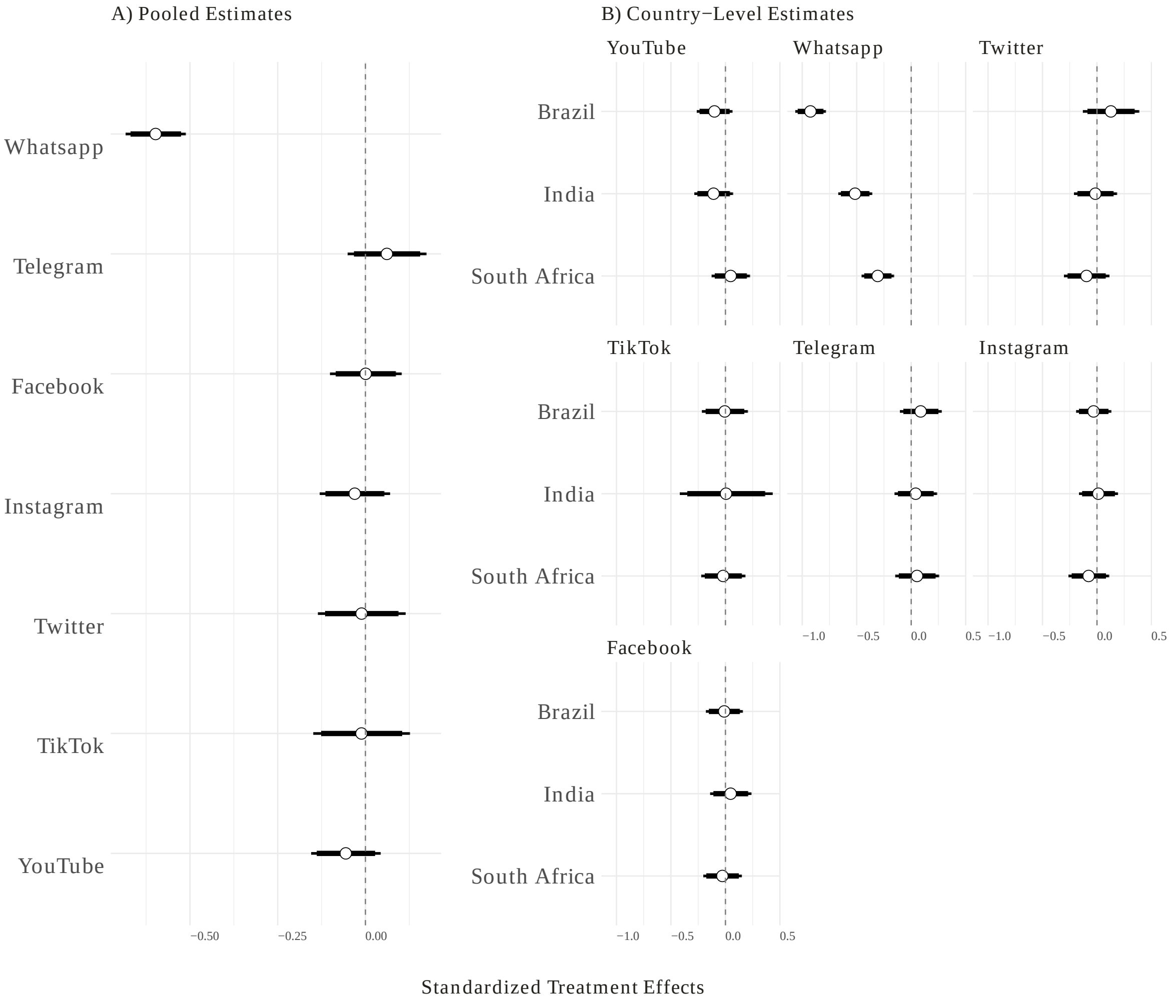
Figure D12: Treatment Effects on Political Interest and Voter Turnout



Notes: See Appendix Tables G35 and G36 for full regression results.

D.4 Substitution to Other Social Media Applications

Figure D13: Treatment Effects on Substitution to Distinct Social Media Platforms



Notes: See Appendix Tables G31 and G32 for full regression results.

D.5 Multiple Hypotheses Testing

In this section, we present the results reported in the paper for our pre-registered outcomes after correcting for multiple hypotheses testing. We use the Benjamini-Hochberg sharpened False Discovery Rate (FDR) adjustment. We adjust the information outcomes by six hypotheses and the attitudinal outcomes by four hypotheses. Table D14 presents the results. Except for **H1**, all of our outcomes which have statistically significant effects at conventional 95% confidence intervals remain so with adjusted p-values smaller than 0.05. With adjusted p-values, the treatment effect on exposure to low-quality political discussions is only significant at the 90% level.

Table D14: Unadjusted and FDR Adjusted P-Values Testing Each Hypothesis)

Hypotheses	Outcome	Unadjusted P-Value	FDR Adjusted P-Value
Information Outcomes			
H1	Low-Quality Political Discussions	0.049	0.074
H2	Online Incivility	0.022	0.044
H3a	Misinformation Recall	0.000	0.001
H3b	News Recall	0.000	0.000
H4a	Misinformation Accuracy	0.313	0.376
H4b	News Accuracy	0.648	0.648
Attitudinal Outcomes			
H5	Partisan Polarization	0.885	0.885
H6	Identity-based Prejudice	0.546	0.728
H7	Issue Polarization	0.195	0.635
H8	Candidate Favorability	0.317	0.635

Note: The unadjusted p-values are estimated using multilevel models for the pooled treatment effects with covariates selected via Lasso for precision. For Information Outcomes, we adjust for 6 comparisons, while for the Attitudinal Outcomes, we adjust for 4 simultaneous comparisons. We use Benjamini-Hochberg sharpened False Discovery Rate (FDR) adjustment.

D.6 Heterogeneous Treatment Effects

We present pre-registered heterogeneous treatment effects analyses for our main outcomes based on digital literacy, age, overall WhatsApp usage, WhatsApp usage for news, and exposure to political content on WhatsApp. Results are presented in the regression tables below. Among these moderators, WhatsApp usage produces the most interesting effects; therefore, we further present marginal effects for this moderator in Figure D14, looking at the group of heavier (above the median) users and lighter (below the median) users. Additionally, we explore heterogeneous effects conditional on the strength of partisan identity (Figure D15 and Table D20).

Table D15: Regression Models: Heterogeneous Effects Conditional on Digital Literacy

	Misinformation Exposure	News Exposure	Misinformation Beliefs	News Knowledge	Online Incivility	Low-Quality Political Discussions	Partisan Polarization	Identity-based Prejudice	Issue Polarization	Candidate Favorability	Subjective Well-Being
Treatment	-0.148+	-0.171*	0.127	0.005	-0.011	0.216	-0.179	-0.037	-0.224	-0.226+	0.485+
	(0.078)	(0.085)	(0.080)	(0.087)	(0.128)	(0.182)	(0.153)	(0.144)	(0.219)	(0.121)	(0.259)
Digital Literacy	-0.093***	-0.009	0.138***	0.043+	0.027	0.066	-0.041	-0.025	-0.040	-0.063+	0.051
	(0.022)	(0.024)	(0.022)	(0.025)	(0.036)	(0.051)	(0.044)	(0.040)	(0.062)	(0.034)	(0.072)
Treatment x Digital Literacy	-0.004	-0.028	-0.041	-0.013	-0.062	-0.186**	0.080	-0.017	0.032	0.076	0.006
	(0.030)	(0.033)	(0.031)	(0.034)	(0.050)	(0.071)	(0.060)	(0.056)	(0.085)	(0.047)	(0.101)
Num.Obs.	2222	2222	2222	2220	2222	2214	2220	2222	2222	2121	2137
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table D16: Regression Models: Heterogeneous Effects Conditional on Age

	Misinformation Exposure	News Exposure	Misinformation Beliefs	News Knowledge	Online Incivility	Low-Quality Political Discussions	Partisan Polarization	Identity-based Prejudice	Issue Polarization	Candidate Favorability	Subjective Well-Being
Treatment	-0.281*	-0.319**	0.005	-0.021	0.064	0.037	-0.260	-0.100	-0.284	-0.147	0.690+
	(0.110)	(0.120)	(0.114)	(0.123)	(0.180)	(0.257)	(0.216)	(0.204)	(0.310)	(0.171)	(0.368)
Age	0.002	0.049	0.025	-0.019	-0.249***	0.006	-0.050	-0.074	0.072	0.161***	0.217*
	(0.031)	(0.034)	(0.032)	(0.034)	(0.050)	(0.071)	(0.061)	(0.056)	(0.086)	(0.048)	(0.101)
Treatment x Age	0.050	0.037	0.018	-0.000	-0.087	-0.092	0.107	0.011	0.051	0.034	-0.083
	(0.043)	(0.047)	(0.045)	(0.048)	(0.070)	(0.100)	(0.084)	(0.079)	(0.121)	(0.067)	(0.144)
Num.Obs.	2221	2221	2221	2219	2221	2213	2219	2221	2221	2120	2136
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table D17: Regression Models: Heterogeneous Effects Conditional on WhatsApp Usage

	Misinformation Exposure	News Exposure	Misinformation Beliefs	News Knowledge	Online Incivility	Low-Quality Political Discussions	Partisan Polarization	Identity-based Prejudice	Issue Polarization	Candidate Favorability	Subjective Well-Being
Treatment	0.212+	-0.079	-0.166	-0.418**	0.179	0.334	-0.236	0.024	-0.113	-0.010	0.679
	(0.125)	(0.137)	(0.130)	(0.140)	(0.205)	(0.291)	(0.245)	(0.231)	(0.352)	(0.194)	(0.417)
WhatsApp Usage	0.048*	-0.014	-0.007	-0.105***	0.130***	0.221***	0.000	0.039	-0.000	0.022	-0.018
	(0.024)	(0.026)	(0.025)	(0.026)	(0.038)	(0.055)	(0.047)	(0.044)	(0.066)	(0.037)	(0.078)
Treatment x WhatsApp Usage	-0.103**	-0.042	0.059+	0.110**	-0.089+	-0.142+	0.062	-0.027	-0.013	-0.016	-0.049
	(0.032)	(0.035)	(0.033)	(0.036)	(0.053)	(0.075)	(0.063)	(0.059)	(0.090)	(0.050)	(0.107)
Num.Obs.	2222	2222	2222	2220	2222	2214	2220	2222	2222	2121	2137
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table D18: Regression Models: Heterogeneous Effects Conditional on WhatsApp Usage for News

	Misinformation Exposure	News Exposure	Misinformation Beliefs	News Knowledge	Online Incivility	Low-Quality Political Discussions	Partisan Polarization	Identity-based Prejudice	Issue Polarization	Candidate Favorability	Subjective Well-Being
Treatment	0.142	-0.105	-0.094	-0.248	0.564+	0.711	-0.031	-0.054	-0.309	-0.271	0.645
	(0.195)	(0.212)	(0.201)	(0.218)	(0.318)	(0.451)	(0.383)	(0.359)	(0.546)	(0.305)	(0.652)
WhatsApp for News	0.088**	0.092**	-0.055*	-0.006	0.125**	0.202***	0.001	0.002	0.106	-0.058	0.143
	(0.028)	(0.030)	(0.028)	(0.030)	(0.044)	(0.061)	(0.055)	(0.049)	(0.074)	(0.043)	(0.089)
Treatment x WhatsApp for News	-0.057	-0.023	0.027	0.042	-0.132*	-0.167*	0.004	-0.004	0.028	0.038	-0.027
	(0.035)	(0.038)	(0.037)	(0.040)	(0.058)	(0.082)	(0.069)	(0.065)	(0.099)	(0.055)	(0.118)
Num.Obs.	2222	2222	2222	2220	2222	2214	2220	2222	2222	2121	2137
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table D19: Regression Models: Heterogeneous Effects Conditional on How Often One Receives Political Content on WhatsApp

	Misinformation Exposure	News Exposure	Misinformation Beliefs	News Knowledge	Online Incivility	Low-Quality Political Discussions	Partisan Polarization	Identity-based Prejudice	Issue Polarization	Candidate Favorability	Subjective Well-Being
Treatment	0.016	-0.197	-0.122	-0.308+	0.482+	0.481	0.210	0.121	-0.072	0.050	0.442
	(0.167)	(0.182)	(0.172)	(0.186)	(0.271)	(0.386)	(0.326)	(0.308)	(0.468)	(0.263)	(0.556)
WhatsApp Political Content	0.065**	0.074**	-0.063*	0.012	0.185***	0.259***	0.047	0.042	0.128+	0.044	-0.024
	(0.025)	(0.027)	(0.025)	(0.027)	(0.039)	(0.056)	(0.049)	(0.045)	(0.068)	(0.038)	(0.079)
Treatment x WhatsApp Political Content	-0.038	-0.008	0.036	0.060	-0.133*	-0.141+	-0.046	-0.041	-0.020	-0.025	0.013
	(0.034)	(0.037)	(0.035)	(0.038)	(0.055)	(0.078)	(0.066)	(0.062)	(0.095)	(0.053)	(0.112)
Num.Obs.	2222	2222	2222	2220	2222	2214	2220	2222	2222	2121	2137
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

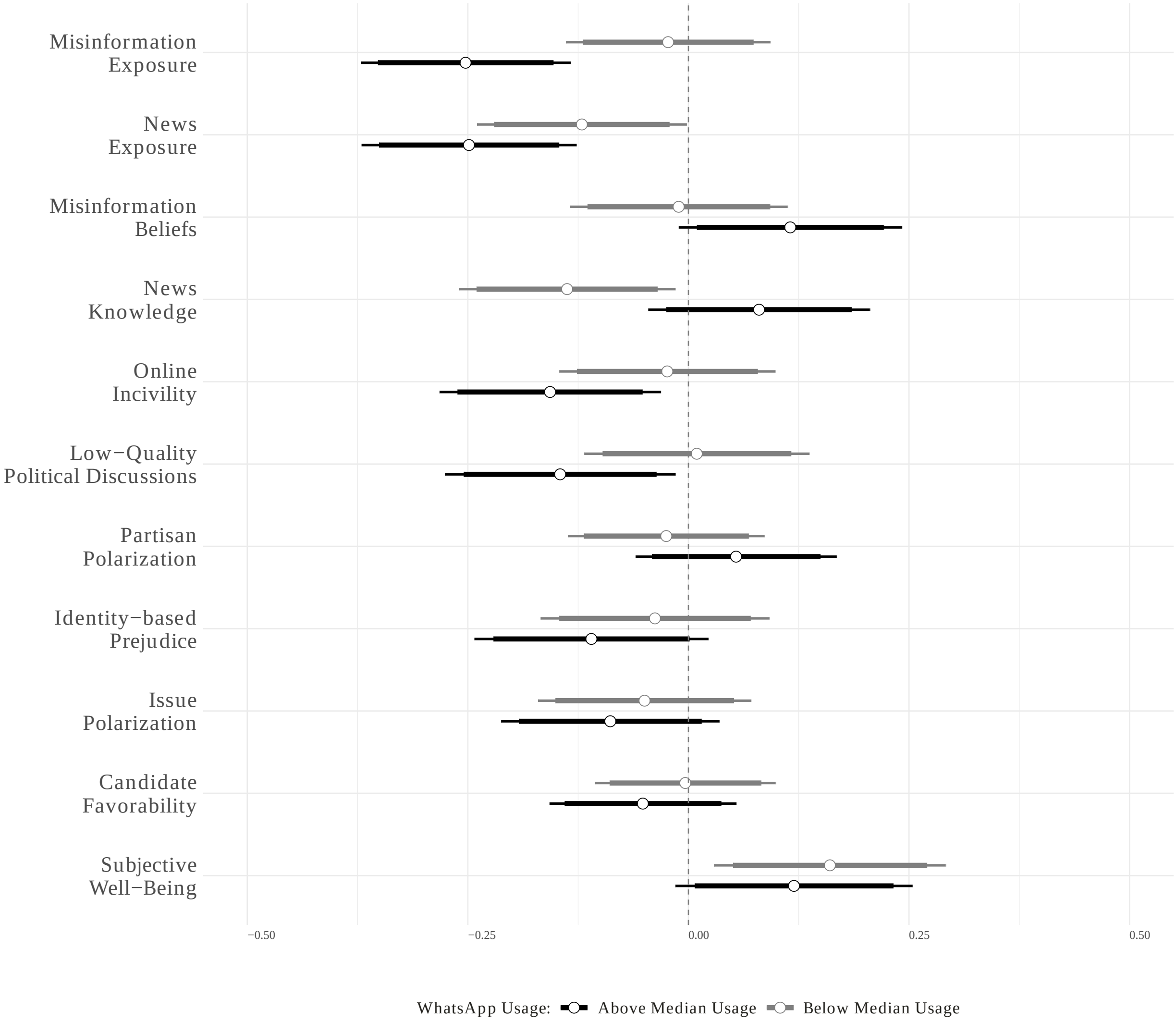
Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table D20: Regression Models: Heterogeneous Effects Conditional on the Salience of Political Identity

	Misinformation Exposure	News Exposure	Misinformation Beliefs	News Knowledge	Online Incivility	Low-Quality Political Discussions	Partisan Polarization	Identity-based Prejudice	Issue Polarization	Candidate Favorability	Subjective Well-Being
Treatment	-0.169**	-0.163*	0.071	0.070	-0.050	-0.072	0.038	-0.048	-0.428*	-0.132	0.497*
	(0.064)	(0.068)	(0.065)	(0.070)	(0.103)	(0.146)	(0.124)	(0.120)	(0.175)	(0.094)	(0.207)
Salience Political Identity	-0.129+	-0.103	0.085	0.092	0.041	0.230	0.359+	-0.263+	-0.291	-0.193+	-0.419+
	(0.074)	(0.113)	(0.072)	(0.079)	(0.170)	(0.241)	(0.205)	(0.138)	(0.290)	(0.112)	(0.226)
Treatment x Salience Political Identity	0.012	-0.137	-0.035	-0.200+	-0.169	-0.330	-0.140	-0.141	0.552*	-0.005	-0.124
	(0.098)	(0.105)	(0.099)	(0.106)	(0.158)	(0.224)	(0.190)	(0.184)	(0.269)	(0.146)	(0.316)
Num.Obs.	1991	1991	1991	1991	1991	1985	1991	1991	1991	1907	1912
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

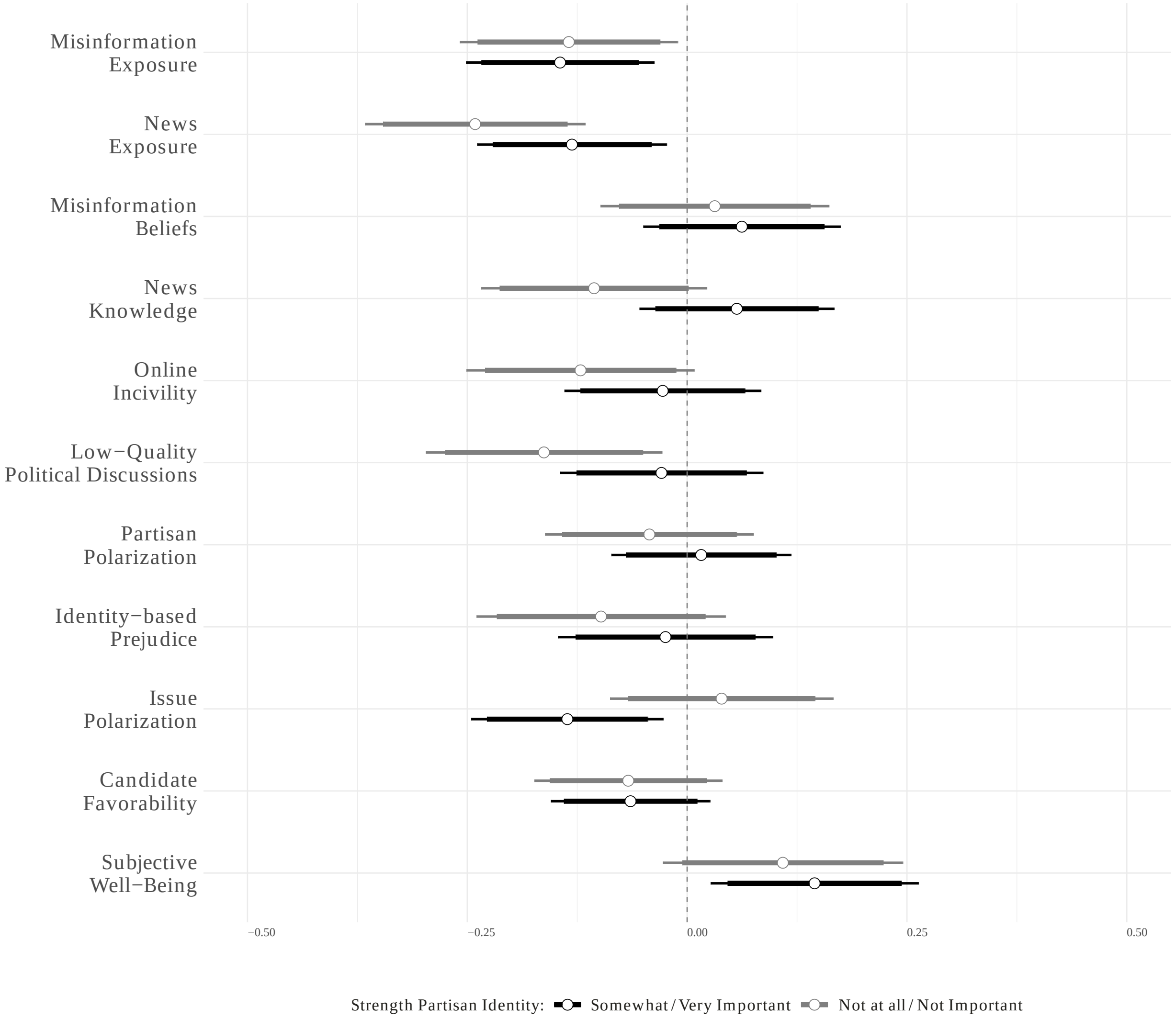
Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Figure D14: Treatment Effects Conditional on Low and High WhatsApp Usage



Notes: See Appendix Table D17 for full regression results.

Figure D15: Treatment Effects Conditional on the Strength of Partisan Identity



Notes: See Appendix Table D20 for full regression results.

D.7 Compliance Estimates and Complier Average Treatment Results

We incentivized users in the treatment condition to alter their WhatsApp usage, either by limiting the amount of time they spend on the app or by refraining from consuming multimedia content on the app. Did our intervention actually lead to a reduction in WhatsApp usage/consumption? The pre-registered models presented in the main text use ITT results, but in this section, we present estimates of Complier Average Treatment Effects (CACE). In summary, we detect substantial compliance with our intervention, and our inferences do not change substantially when comparing

ITT and CACE estimates.

To examine whether our intervention reduced WhatsApp usage and consumption, at the end of every week of the experiment period, we collected a screenshot showing one's weekly WhatsApp screen time from participants in the *Time* condition and a screenshot showing one's cumulative WhatsApp data consumption in the *Multimedia* condition. We use these screenshots to construct a binary indicator of "low WhatsApp usage" which serves as an indicator of compliance for those in the treatment condition and a benchmark to compare against that in the control condition.

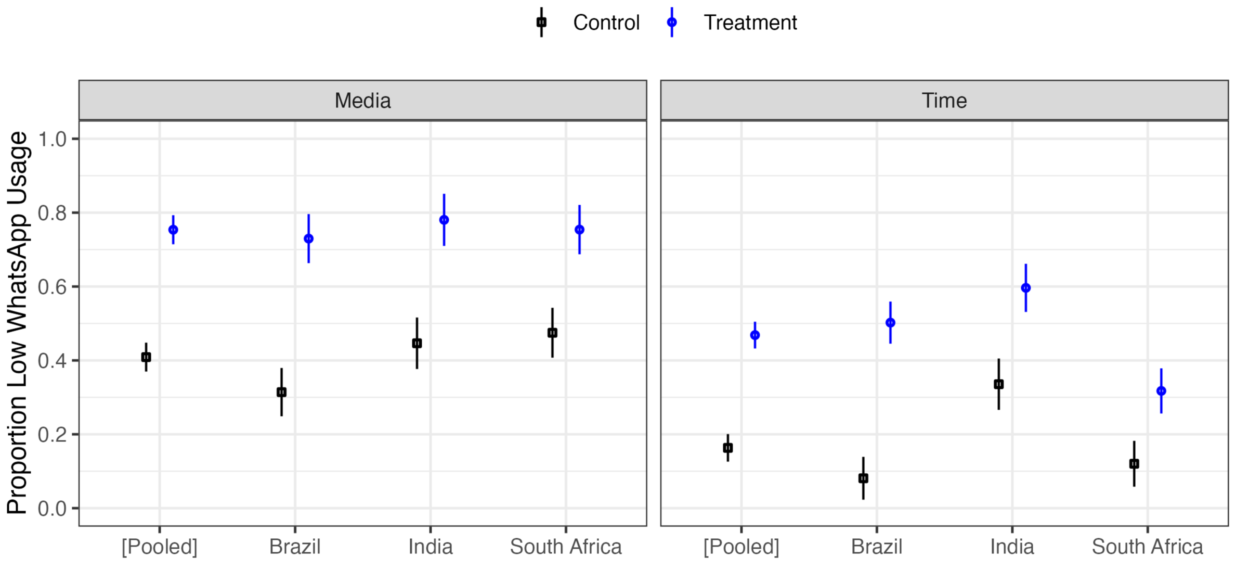
In the *Time* condition, a user is classified as having low WhatsApp usage if each of their last three weekly screenshots shows a daily average WhatsApp screen time of less than 10 minutes. In the *Multimedia* condition, we define low WhatsApp usage as the gap in cumulative megabytes downloaded between the baseline and endline screenshots being lower than the average gap in the control condition for their respective country. A user in the *Multimedia* control condition is classified as such if this gap is lower than 100 megabytes, which we determined to be a reasonable threshold following pilot studies.

Note that if a user submitted a screenshot that was doctored, a duplicate of a previous week's submission, from a different phone as their previous submissions, or indicated their data consumption statistics had been manually reset during the experiment period, we classify them as having violated study rules and mark their screenshots as missing. In practice, about 4% of our sample either violated one of these rules or simply did not submit screenshots; country or treatment status does not predict violations or missingness.

Figure D16 shows the proportion of treatment (in blue) and control (in black) participants classified as having low WhatsApp usage in the *Multimedia* arm (first panel) and in the *Time* arm (second panel). First, we observe considerable variation in control group usage levels across countries, with Brazil having relatively fewer low usage respondents (and relatedly larger treatment effects on reducing usage). Second, across countries and arms, the proportion of participants with low WhatsApp usage is significantly higher in the treatment condition than in the control condition, showing that treatment assignment worked as intended in creating two distinct groups with substantially different levels of time spent on WhatsApp or multimedia content consumed during elections. Third, we note that under the strictest definitions of compliance, there is a stronger degree of compliance in the *Multimedia* arm compared to the *Time* arm, suggesting that, given the same monetary incentives, duration, and baseline usage levels, individuals are more willing and able to stop consuming multimedia content on WhatsApp compared to limiting WhatsApp usage overall to 10 minutes per day.

That said, though we observe that the proportion of *Time* treatment respondents who spent the *entire* experiment period under the 10-minute daily limit was about 0.47 (as shown in Figure D16), the proportion who never exceeded 30 minutes was over 0.70. Further, in any given week, the proportion who stayed under 10 minutes was over 0.60. Comparing usage in the pre-treatment week with the average during-treatment week, we also observe a reduction of about 0.84 SD in WhatsApp screen time within the treatment condition. These numbers confirm that the *Time* treatment was quite effective in spurring respondents to limit their usage.

Figure D16: Treatment Assignment and Low WhatsApp Usage During Experiment



We use this binary ”low dosage usage” measure to estimate Compliance Average Treatment Effects, using an instrumental variable estimator. The following table presents the pooled results. As expected, we find much larger effects among the sample of compliers, but our inferences do not change substantially when compared to the primary results of the paper.

Table D21: Regression Models: Compliance Average Treatment Effects on Primary Outcomes.

	Misinformation Exposure	News Exposure	Misinformation Beliefs	News Knowledge	Online Incivility	Low-Quality Political Discussions	Partisan Polarization	Identity-based Prejudice	Issue Polarization	Candidate Favorability	Subjective Well-Being
Intercept	0.556*	1.021*	2.372***	4.465***	-1.150	-1.191***	-2.841*	3.974***	2.680+	1.404	-1.567**
	(0.243)	(0.501)	(0.192)	(0.503)	(1.389)	(0.265)	(1.200)	(0.805)	(1.560)	(0.932)	(0.531)
CACE estimate	-0.519***	-0.734***	0.144	-0.066	-0.598+	-0.564*	-0.044	-0.190	-0.442	-0.216	1.496**
	(0.143)	(0.157)	(0.149)	(0.160)	(0.322)	(0.229)	(0.278)	(0.242)	(0.398)	(0.223)	(0.460)
Num.Obs.	2222	2222	2222	2220	2214	2222	2222	2220	2222	2121	2137
R2 Marg.	0.071	0.114	0.028	0.054	0.100	0.109	0.269	0.166	0.080	0.299	0.019
R2 Cond.	0.136	0.211	0.048	0.064		0.111	0.278	0.169	0.081	0.348	0.021
RMSE	1.05	1.14	1.09	1.17	2.36	1.70	2.04	1.78	2.95	1.59	3.43
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Country Random Intercepts	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

D.8 Baseline Difference for Primary Outcomes Across Countries

Figures D17, D18, and D19 present average responses to key post-treatment and pre-treatment variables among untreated respondents in each country. In the main text, we consider how cross-country differences in baseline attitudes and exposure to content can help inform differences in treatment effects across countries. We also note that there are no discernible differences between our "time control group" and "media control group" with respect to these variables, which alleviates any concerns one may have about pooling the control groups in our analyses.

Figure D17: Control group responses by country (headline recall and accuracy outcomes)

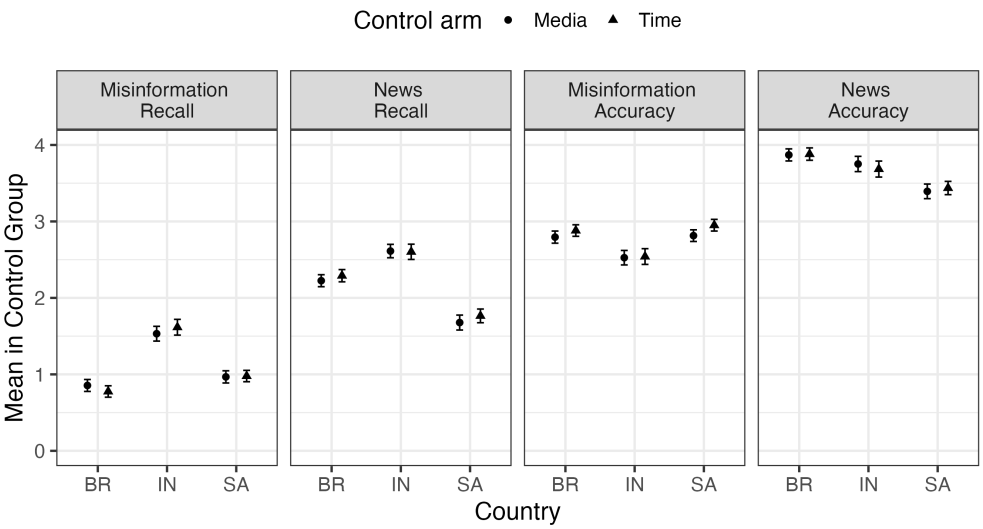


Figure D18: Control group responses by country (incivility and quality of discussions outcomes)

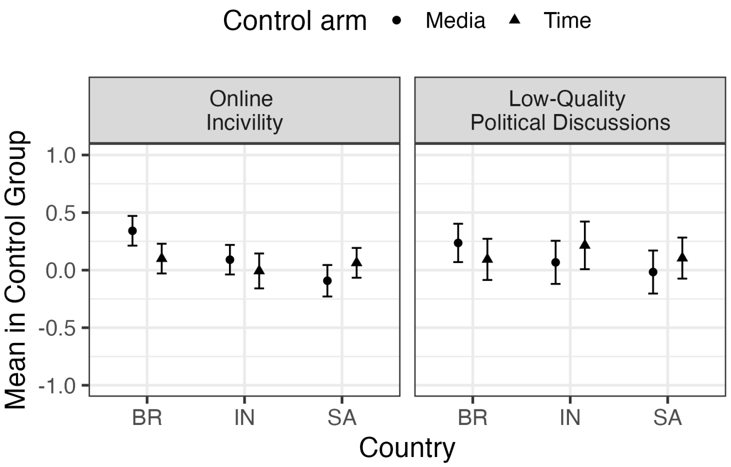
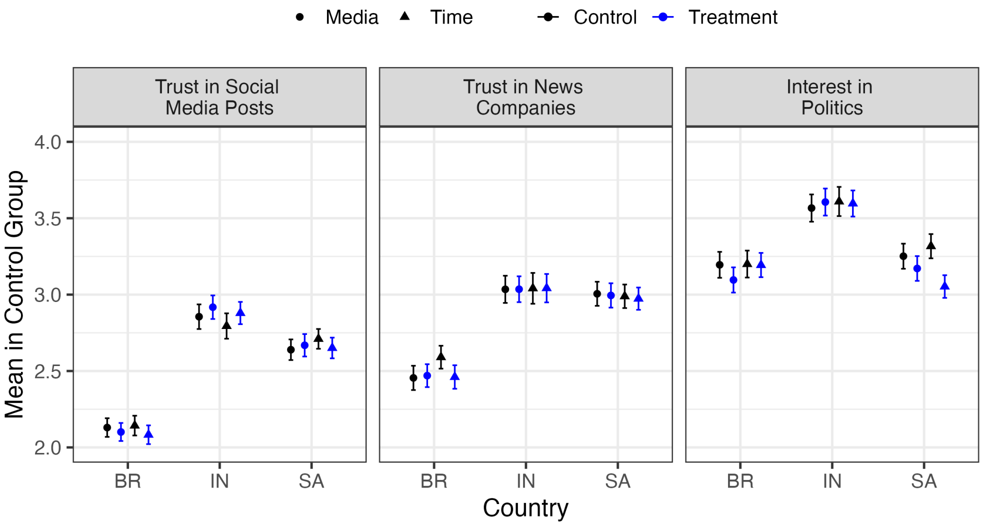


Figure D19: Trust and interest by country (pre-treatment variables)

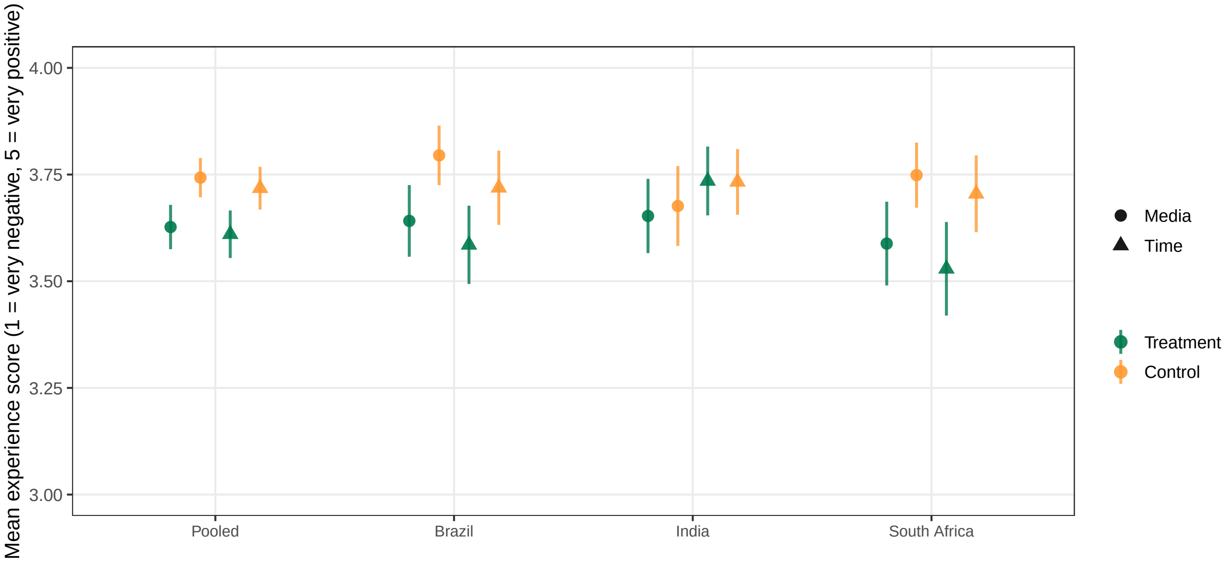


E Participant Perspectives on Study Participation

In this section, we explore respondents' perspectives on participating in the study. We use data from two questions in the endline survey: a close-ended question on overall participation experience (sent to all respondents) and an open-ended question on the experience of limiting WhatsApp usage for four weeks (sent only to participants assigned to the treatment condition).

First, Figure E20 presents responses to the close-ended question, plotting the mean response by country and treatment condition. Across the different countries and conditions, the average response was above 3.5 (where 3 indicates neutral and 4 indicates positive), though treated individuals did report having a marginally less positive experience than those in the control—which is not surprising given the major lifestyle change the former undertook compared to the latter's experience of just completing a few short surveys. These results provide a promising foundation when considering the scalability of such interventions. Though the intervention may have been disruptive, respondents generally did not find it to be unacceptable. As we discuss next, their open-ended responses corroborate this argument and reveal a rather nuanced picture, highlighting that the experience was rewarding and offered unexpected benefits for many participants even if it brought some difficulties with it.

Figure E20: Close-Ended Study Experience Responses, By Country and Condition



Each treated participant was further asked to narrate their experience reducing WhatsApp usage during the preceding four weeks.[23] Between 98% and 99.5% of respondents across the three countries and two treatment types replied. In reading their responses, we observed the following themes come up repeatedly: the difficulty of refraining from using an app as deeply embedded in individuals' lives as WhatsApp is, the benefits of not seeing unnecessary or annoying content anymore as a function of having to prioritize what one views during their limited time on the app, and the advantage of suddenly having significantly more time to pursue other activities. For example, a South African participant wrote, "it was hard at first and felt like i was missing out on the rest of the day's interactions but i got used to it and started feeling like i had so much time to my hands and could do a lot more things." Similarly, a Brazilian user shared, "It was a big challenge, the first few days were the hardest, but I kept myself busy with other things, my boyfriend liked that I spoke to him more in person instead of on WhatsApp." In India, another respondent noted, "a lot of negative energy has lifted, i really felt this was a great move because whenever i used to consume whatsapp media most of it will be false news and negativity... since i haven't consumed it for a long time, i feel very positive i never came across anything offensive or false."

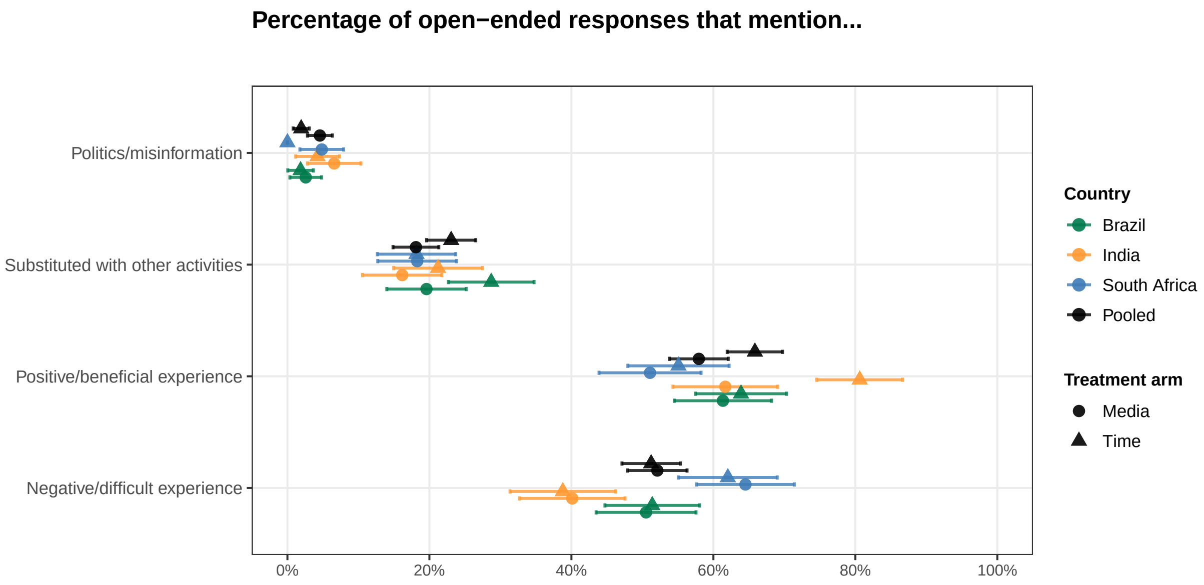
Accordingly, we used Claude-3.5-Sonnet, a large language model, to systematically classify the text responses across four dimensions. First, whether the respondent mentioned positive facets of the experience, such as enjoyment, ease of reducing WhatsApp usage, reduced anxiety or stress as a result of being away from social media, or improved personal well-being. Second, whether the respondent reported increased anxiety, difficulty abstaining, missing out on important or interesting messages, or other negative feelings. Third, whether the respondent mentioned substituting WhatsApp with other activities like spending time with friends and family, watching television, pursuing hobbies, being more productive at work, or using other social media apps. Fourth, whether they explicitly mentioned politics, elections, or misinformation. Figure E21 shows the proportion of open-ended responses coded as "yes" in each category across the entire sample and

[23] We did not ask control condition participants this question as we were specifically interested in what individuals thought about the usage reduction experience, rather than how the overall study experience varied by the type of experience.

by country, disaggregated by treatment type.

We observe that respondents often had both positive and negative takeaways. Overall, pooling across the three countries, we find that 64% of respondents reported positive experiences and 52% reported negative experiences, suggesting that many participants had mixed feelings about reducing WhatsApp usage (and felt comfortable sharing these feelings candidly with the research team). Further, 21% mentioned substitution activities and 3% explicitly referenced politics or misinformation—without anybody being prompted to write about any of these themes.[24]

Figure E21: Classification of Open-Ended Study Experience Responses, By Country and Condition



[24] Disaggregating by country, we find that in Brazil, 63% of the respondents wrote about positive experiences while 51% mentioned negative experiences. Approximately 24% mentioned substituting WhatsApp with other activities, while only 2% explicitly referenced politics or misinformation. India showed the highest proportion of positive experiences at 71%, with 39% reporting negative experiences and 19% mentioning substitution activities. Political content was mentioned by 5% of Indian respondents, the highest rate among the three countries. We uncovered a somewhat different pattern in South Africa, where 53% mentioned positive experiences but 63% brought up negative experiences.

F Unadjusted ITT Results

In this section, we present non-adjusted ITT results (without any covariates) for robustness. All models reported in the paper use, as pre-registered, ITT results with covariates selected via LASSO procedure.

Table F22: Treatment Effects: Intention-to-Treat Unadjusted Models

Outcomes	Pooled Effects	Brazil	India	South Africa
Information Outcomes				
Misinformation Recall	-0.142 (0.039) ⋆⋆⋆	-0.14 (0.065) ⋆	-0.183 (0.073) ⋆	-0.109 (0.068)
News Recall	-0.196 (0.041) ⋆⋆⋆	-0.295 (0.068) ⋆⋆⋆	-0.042 (0.076)	-0.221 (0.072) ⋆⋆
Misinformation Accuracy	0.036 (0.041)	0.046 (0.068)	0.081 (0.076)	-0.016 (0.072)
News Accuracy	-0.029 (0.042)	-0.136 (0.068) ⋆	0.011 (0.077)	0.054 (0.072)
Online Toxicity	-0.094 (0.042) ⋆	-0.222 (0.069) ⋆⋆	-0.055 (0.078)	0.016 (0.073)
Low-Quality Political Discussions	-0.081 (0.043) ○	-0.117 (0.07) ○	-0.09 (0.079)	-0.032 (0.074)
Attitudinal Outcomes				
Partisan Polarization	-0.032 (0.043)	-0.072 (0.07)	-0.116 (0.078)	0.087 (0.074)
Identity-based Prejudice	-0.042 (0.043)	-0.042 (0.07)	-0.148 (0.079) ○	0.053 (0.074)
Issue Polarization	-0.065 (0.041)	-0.11 (0.068)	-0.004 (0.076)	-0.07 (0.072)
Candidate Favorability	-0.06 (0.04)	-0.103 (0.066)	-0.06 (0.073)	-0.009 (0.071)
Subjective Well-Being and Substitution				
Watching TV	0.091 (0.042) ⋆	0.124 (0.068) ○	0.041 (0.077)	0.098 (0.072)
Time with Friends	0.06 (0.043)	0.217 (0.07) ⋆⋆	-0.046 (0.079)	-0.02 (0.074)
Hobbies	0.22 (0.042) ⋆⋆⋆	0.323 (0.068) ⋆⋆⋆	0.179 (0.077) ⋆	0.141 (0.072) ○
Other Social Media Apps	-0.106 (0.043) ⋆	-0.136 (0.07) ○	-0.159 (0.079) ⋆	-0.026 (0.074)
Subjective Well-Being	0.148 (0.044) ⋆⋆⋆	0.283 (0.072) ⋆⋆⋆	0.064 (0.079)	0.072 (0.076)

Note: Standard errors in parentheses. All models use multilevel estimations with random intercepts by country. ○ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

G Regression Tables

In this section, we present numerical results for all ITT estimates reported in the main manuscript and the appendix. All models include covariates selected via LASSO for each specific outcome. Given that our inference is based on random assignment across participants, and control variables serve the sole purpose of increasing the precision of our estimates, we do not present the results for each control variable in the models below.

Table G23: Full Regression Models (Information and Political Outcomes, Pooled Treatments)

	Misinformation Exposure	News Exposure	Misinformation Beliefs	News Knowledge	Online Incivility	Low-Quality Political Discussions	Partisan Polarization	Identity-based Prejudice	Issue Polarization	Candidate Favorability
Intercept	0.512*	0.870+	2.395***	4.465***	-1.300***	-1.185	-2.848*	3.951***	2.614+	1.387
	(0.238)	(0.493)	(0.186)	(0.503)	(0.259)	(1.388)	(1.197)	(0.806)	(1.553)	(0.931)
Treatment Pooled	-0.163***	-0.230***	0.048	-0.020	-0.168*	-0.191+	-0.014	-0.048	-0.159	-0.068
	(0.045)	(0.049)	(0.046)	(0.050)	(0.072)	(0.101)	(0.087)	(0.076)	(0.126)	(0.070)
Num.Obs.	2222	2222	2222	2220	2222	2214	2222	2220	2222	2121
R2 Marg.	0.072	0.117	0.028	0.054	0.109	0.100	0.269	0.166	0.080	0.300
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country RE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G24: Full Regression Models (Information and Political Outcomes, Pooled Treatments, By Country)

	Misinformation Exposure	News Exposure	Misinformation Beliefs	News Knowledge	Online Incivility	Low-Quality Political Discussions	Partisan Polarization	Identity-based Prejudice	Issue Polarization	Candidate Favorability
Intercept	0.347	1.039	2.403***	4.583***	-1.066***	-1.058	-3.088**	3.962***	2.981+	1.768
	(0.375)	(1.033)	(0.209)	(0.512)	(0.255)	(1.400)	(1.190)	(0.798)	(1.582)	(1.189)
Treatment	-0.177*	-0.353***	0.056	-0.141+	-0.413***	-0.289+	-0.007	0.003	-0.252	-0.087
	(0.074)	(0.080)	(0.076)	(0.082)	(0.119)	(0.166)	(0.143)	(0.125)	(0.206)	(0.114)
Country: India	0.529	0.115	-0.205	-0.168	-0.415**	-0.114	0.313+	-0.160	0.064	-0.199
	(0.477)	(1.323)	(0.204)	(0.222)	(0.136)	(0.256)	(0.173)	(0.149)	(0.355)	(1.121)
Country: South Africa	-0.005	-0.553	0.166	-0.379+	-0.423***	-0.388+	0.408*	-0.173	-0.144	-0.876
	(0.475)	(1.322)	(0.202)	(0.218)	(0.127)	(0.227)	(0.162)	(0.142)	(0.343)	(1.120)
Treatment x India	-0.037	0.297*	0.037	0.136	0.279	-0.026	-0.222	-0.259	0.262	-0.010
	(0.111)	(0.120)	(0.114)	(0.124)	(0.178)	(0.251)	(0.215)	(0.187)	(0.310)	(0.170)
Treatment x South Africa	0.074	0.106	-0.059	0.240*	0.489**	0.308	0.175	0.074	0.045	0.069
	(0.107)	(0.117)	(0.111)	(0.119)	(0.172)	(0.242)	(0.208)	(0.181)	(0.300)	(0.168)
Num.Obs.	2222	2222	2222	2220	2222	2214	2222	2220	2222	2121
R2 Marg.	0.126	0.135	0.039	0.068	0.111	0.102	0.268	0.167	0.082	0.321
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country RE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. Brazil is the omitted baseline country in each regression model. Models include a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G25: Full Regression Models (Information and Political Outcomes, Unpooled Treatments)

	Misinformation Exposure	News Exposure	Misinformation Beliefs	News Knowledge	Online Incivility	Low-Quality Political Discussions	Partisan Polarization	Identity-based Prejudice	Issue Polarization	Candidate Favorability
Intercept	0.513*	0.876+	2.394***	4.466***	-1.288***	-1.157	-2.850*	3.948***	2.595+	1.383
	(0.238)	(0.493)	(0.186)	(0.503)	(0.260)	(1.388)	(1.197)	(0.806)	(1.553)	(0.931)
Treatment Media	-0.161**	-0.247***	0.040	-0.031	-0.077	-0.292*	0.002	-0.021	-0.115	-0.058
Treatment Time	-0.166**	-0.213***	0.055	-0.010	-0.256**	-0.092	-0.030	-0.075	-0.201	-0.079
	(0.055)	(0.059)	(0.056)	(0.061)	(0.088)	(0.124)	(0.106)	(0.093)	(0.153)	(0.085)
Num.Obs.	2222	2222	2222	2220	2222	2214	2222	2220	2222	2121
R2 Marg.	0.072	0.117	0.028	0.054	0.110	0.101	0.269	0.166	0.080	0.300
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country RE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G26: Full Regression Models (Information and Political Outcomes, Unpooled Treatments, By Country)

	Misinformation Exposure	News Exposure	Misinformation Beliefs	News Knowledge	Online Incivility	Low-Quality Political Discussions	Partisan Polarization	Identity-based Prejudice	Issue Polarization	Candidate Favorability
Intercept	0.345	1.034	2.403***	4.581***	-1.055***	-1.062	-3.100**	3.968***	2.965+	1.774
	(0.375)	(1.034)	(0.209)	(0.560)	(0.255)	(1.400)	(1.191)	(0.799)	(1.584)	(1.193)
Treatment Media	-0.151+	-0.342***	0.023	-0.163	-0.323*	-0.543**	-0.007	0.053	-0.201	-0.074
	(0.092)	(0.100)	(0.095)	(0.102)	(0.147)	(0.207)	(0.178)	(0.155)	(0.257)	(0.141)
Treatment Time	-0.201*	-0.362***	0.087	-0.121	-0.496***	-0.056	-0.006	-0.042	-0.298	-0.099
	(0.089)	(0.097)	(0.092)	(0.099)	(0.143)	(0.201)	(0.172)	(0.150)	(0.249)	(0.138)
Treatment Media x India	-0.101	0.216	0.091	0.143	0.216	0.119	-0.252	-0.332	0.262	0.005
	(0.136)	(0.148)	(0.141)	(0.152)	(0.219)	(0.308)	(0.264)	(0.231)	(0.381)	(0.208)
Treatment Time x India	0.025	0.377**	-0.014	0.132	0.334	-0.146	-0.190	-0.190	0.257	-0.027
	(0.135)	(0.146)	(0.139)	(0.150)	(0.217)	(0.305)	(0.261)	(0.228)	(0.377)	(0.207)
Treatment Media x SA	0.061	0.091	-0.031	0.267+	0.543*	0.626*	0.253	0.076	0.019	0.048
	(0.132)	(0.144)	(0.136)	(0.147)	(0.213)	(0.298)	(0.256)	(0.223)	(0.370)	(0.206)
Treatment Time x SA	0.085	0.119	-0.084	0.216	0.428*	0.009	0.097	0.068	0.067	0.090
	(0.130)	(0.142)	(0.134)	(0.145)	(0.210)	(0.295)	(0.253)	(0.220)	(0.365)	(0.205)
Num.Obs.	2222	2222	2222	2220	2222	2214	2222	2220	2222	2121
R2 Marg.	0.126	0.135	0.040	0.065	0.113	0.104	0.268	0.167	0.082	0.320
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country RE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Robust standard errors in parentheses. Brazil is the omitted baseline country in each regression model. Models include a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G27: Full Regression Models (Non-Political Outcomes, Pooled Treatments)

	Subjective Well-Being	Watching TV	Time with Friends	Other Social Media Apps	Hobbies
Intercept	-1.232*	2.898***	2.835***	3.492***	2.940***
	(0.498)	(0.196)	(0.167)	(0.184)	(0.182)
Treatment Pooled	0.501***	0.128*	0.075	-0.135*	0.262***
	(0.149)	(0.054)	(0.049)	(0.055)	(0.048)
Num.Obs.	2137	2222	2222	2222	2222
R2 Marg.	0.018	0.020	0.020	0.009	0.029
Controls	Yes	Yes	Yes	Yes	Yes
Country RE	Yes	Yes	Yes	Yes	Yes

Notes: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G28: Full Regression Models (Non-Political Outcomes, Pooled Treatments, By Country)

	Subjective Well-Being	Watching TV	Time with Friends	Other Social Media Apps	Hobbies
Intercept	-1.508**	2.837***	2.776***	3.532***	2.741***
	(0.510)	(0.187)	(0.176)	(0.194)	(0.278)
Treatment	0.982***	0.172+	0.257**	-0.172+	0.381***
	(0.245)	(0.089)	(0.081)	(0.090)	(0.079)
Country: India	0.192	0.266*	0.183	0.065	0.395
	(0.274)	(0.104)	(0.116)	(0.126)	(0.330)
Country: South Africa	0.506+	-0.053	0.035	-0.163	0.244
	(0.259)	(0.098)	(0.111)	(0.121)	(0.328)
Treatment x India	-0.790*	-0.115	-0.307*	-0.029	-0.168
	(0.364)	(0.134)	(0.122)	(0.135)	(0.119)
Treatment x South Africa	-0.735*	-0.030	-0.276*	0.136	-0.208+
	(0.357)	(0.129)	(0.117)	(0.130)	(0.115)
Num.Obs.	2137	2222	2222	2222	2222
R2 Marg.	0.022	0.023	0.026	0.011	0.042
Controls	Yes	Yes	Yes	Yes	Yes
Country RE	Yes	Yes	Yes	Yes	Yes

Notes: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. Brazil is the omitted baseline country in each regression model. Models include a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G29: Full Regression Models (Non-Political Outcomes, Unpooled Treatments)

	Subjective Well-Being	Watching TV	Time with Friends	Other Social Media Apps	Hobbies
Intercept	-1.246*	2.901***	2.822***	3.486***	2.937***
	(0.498)	(0.196)	(0.167)	(0.184)	(0.182)
Treatment Media	0.389*	0.146*	-0.005	-0.169*	0.239***
	(0.182)	(0.067)	(0.061)	(0.067)	(0.059)
Treatment Time	0.612***	0.110+	0.152*	-0.102	0.285***
	(0.182)	(0.066)	(0.060)	(0.067)	(0.059)
Num.Obs.	2137	2222	2222	2222	2222
R2 Marg.	0.019	0.020	0.023	0.009	0.029
Controls	Yes	Yes	Yes	Yes	Yes
Country RE	Yes	Yes	Yes	Yes	Yes

Notes: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G30: Full Regression Models (Non-Political Outcomes, Unpooled Treatments, By Country)

	Subjective Well-Being	Watching TV	Time with Friends	Other Social Media Apps	Hobbies
Intercept	-1.540**	2.836***	2.759***	3.530***	2.735***
	(0.510)	(0.228)	(0.176)	(0.194)	(0.262)
Treatment Media	1.079***	0.270*	0.253*	-0.255*	0.377***
	(0.303)	(0.111)	(0.100)	(0.112)	(0.098)
Treatment Time	0.892**	0.082	0.261**	-0.097	0.385***
	(0.296)	(0.107)	(0.097)	(0.108)	(0.095)
Treatmet Media x India	-1.329**	-0.217	-0.485**	0.078	-0.321*
	(0.446)	(0.165)	(0.149)	(0.166)	(0.146)
Treatment Time x India	-0.250	-0.022	-0.124	-0.130	-0.011
	(0.443)	(0.163)	(0.148)	(0.164)	(0.144)
Treatment Media x SA	-0.846+	-0.172	-0.331*	0.184	-0.124
	(0.437)	(0.160)	(0.145)	(0.161)	(0.141)
Treatment Time x SA	-0.627	0.103	-0.220	0.096	-0.291*
	(0.435)	(0.157)	(0.142)	(0.158)	(0.139)
Num.Obs.	2137	2222	2222	2222	2222
R2 Marg.	0.025	0.024	0.030	0.012	0.045
Controls	Yes	Yes	Yes	Yes	Yes
Country RE	Yes	Yes	Yes	Yes	Yes

Notes: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. Brazil is the omitted baseline country in each regression model. Models include a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G31: Full Regression Models (Social Media Substitution, Pooled Treatments)

	WhatsApp	Telegram	Facebook	Instagram	Twitter	TikTok	YouTube
Intercept	3.110***	2.632***	3.408***	3.389***	2.355***	3.219***	2.977***
	(0.198)	(0.243)	(0.243)	(0.358)	(0.348)	(0.362)	(0.326)
Treatment Pooled	-0.810***	0.071	0.001	-0.035	-0.013	-0.013	-0.065
	(0.059)	(0.066)	(0.060)	(0.059)	(0.073)	(0.081)	(0.058)
Num.Obs.	2124	1671	2040	1938	1234	1283	2066
R2 Marg.	0.091	0.023	0.003	0.017	0.008	0.018	0.012
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country RE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G32: Full Regression Models (Social Media Substitution, Pooled Treatments, By Country)

	WhatsApp	Telegram	Facebook	Instagram	Twitter	TikTok	YouTube
Intercept	3.295***	2.687***	3.201***	3.634***	1.797*	3.370***	2.879***
	(0.201)	(0.234)	(0.287)	(0.249)	(0.833)	(0.640)	(0.637)
Treatment	-1.253***	0.100	-0.013	-0.036	0.148	-0.007	-0.117
	(0.096)	(0.112)	(0.099)	(0.094)	(0.151)	(0.123)	(0.096)
Country: India	-0.320**	0.099	0.157	0.090	0.839	-0.639	0.632
	(0.111)	(0.135)	(0.299)	(0.234)	(1.119)	(0.831)	(0.862)
Country: South Africa	-0.320**	-0.251+	0.473	-0.814***	0.783	0.104	-0.349
	(0.105)	(0.132)	(0.297)	(0.233)	(1.119)	(0.816)	(0.861)
Treatment x India	0.555***	-0.053	0.068	0.051	-0.165	0.013	-0.010
	(0.143)	(0.160)	(0.148)	(0.140)	(0.190)	(0.279)	(0.141)
Treatment x South Africa	0.836***	-0.039	-0.020	-0.052	-0.259	-0.017	0.172
	(0.139)	(0.162)	(0.143)	(0.144)	(0.194)	(0.171)	(0.140)
Num.Obs.	2124	1671	2040	1938	1234	1283	2066
R2 Marg.	0.109	0.037	0.023	0.102	0.049	0.037	0.077
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country RE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Robust standard errors in parentheses. Brazil is the omitted baseline country in each regression model. Models include a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G33: Full Regression Models (Self-Reported Exposure to News, Pooled Treatments)

	False News	True News	Political News	Election-Violence News	Sports News
Intercept	3.124***	3.128***	3.294***	3.199***	3.055***
	(0.151)	(0.132)	(0.189)	(0.163)	(0.179)
Treatment	-0.113**	-0.027	-0.041	-0.084+	-0.053
	(0.039)	(0.037)	(0.041)	(0.047)	(0.042)
Num.Obs.	2222	2222	2222	2222	2222
R2 Marg.	0.014	0.014	0.018	0.016	0.040
Controls	Yes	Yes	Yes	Yes	Yes
Country RE	Yes	Yes	Yes	Yes	Yes

Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G34: Full Regression Models (Self-Reported Exposure to News, Pooled Treatments, By Country)

	False News	True News	Political News	Election-Violence News	Sports News
Intercept	3.110***	3.078***	3.142***	3.110***	2.893***
	(0.137)	(0.198)	(0.339)	(0.177)	(0.425)
Treatment	-0.145*	-0.031	-0.110+	-0.090	-0.158*
	(0.065)	(0.061)	(0.067)	(0.078)	(0.070)
Country: India	-0.101	0.193	0.050	0.187	0.202
	(0.082)	(0.226)	(0.444)	(0.135)	(0.571)
Country: South Africa	0.123	-0.025	0.401	0.097	0.287
	(0.078)	(0.224)	(0.443)	(0.131)	(0.570)
Treatment x India	0.001	-0.071	0.135	-0.084	0.204+
	(0.097)	(0.092)	(0.101)	(0.117)	(0.105)
Treatment x South Africa	0.094	0.074	0.087	0.094	0.135
	(0.094)	(0.088)	(0.097)	(0.113)	(0.101)
Num.Obs.	2222	2222	2222	2222	2222
R2 Marg.	0.028	0.023	0.047	0.021	0.056
Controls	Yes	Yes	Yes	Yes	Yes
Country RE	Yes	Yes	Yes	Yes	Yes

Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Robust standard errors in parentheses. Brazil is the omitted baseline country in each regression model. Models include a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G35: Full Regression Models (Additional Political Outcomes, Pooled Treatments)

	Interest in Politics	Voter Turnout
Intercept	1.690***	0.670***
	(0.203)	(0.053)
Treatment Pooled	-0.055	0.007
	(0.051)	(0.013)
Num.Obs.	2222	2222
R2 Marg.	0.056	0.014
Controls	Yes	Yes
Country RE	Yes	Yes

Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Robust standard errors in parentheses. All models employ multilevel estimation with random intercepts at the country level, along with a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

Table G36: Full Regression Models (Additional Political Outcomes, Pooled Treatments, By Country)

	Interest in Politics	Voter Turnout
Intercept	1.537***	0.695***
	(0.246)	(0.052)
Treatment	-0.101	-0.006
	(0.083)	(0.022)
Country: India	0.320	0.015
	(0.264)	(0.044)
Country: South Africa	0.170	-0.088*
	(0.262)	(0.043)
Treatment x India	0.160	0.021
	(0.125)	(0.032)
Treatment x South Africa	-0.004	0.022
	(0.121)	(0.031)
Num.Obs.	2222	2222
R2 Marg.	0.073	0.033
Controls	Yes	Yes
Country RE	Yes	Yes

Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Robust standard errors in parentheses. Brazil is the omitted baseline country in each regression model. Models include a list of covariates selected via Lasso for each outcome. Covariate estimates are not listed as they were included only to increase precision and do not carry substantive meaning.

H Deviations from the Pre-Analysis Plan

We made minor deviations from the Pre-Analysis Plan (PAP) [LINK REDACTED]. First, our manuscript reorders the hypotheses in the PAP. In the PAP, our first hypothesis relates to the information exposure outcomes; these hypotheses are now H3a and H3b in the manuscript. Hypotheses 5 and 6 in the PAP are now H1 and H2 in the manuscript. The order of the other hypotheses is updated accordingly. Second, we slightly edited the wording of the hypotheses to improve readability. Third, in the PAP, we categorized hypotheses as either primary or secondary, which we do not do in the manuscript. This decision has no impact on our pre-registered multiple hypotheses adjustment, as we perform the pre-registered adjustment for all eight hypotheses, encompassing both the primary and the secondary hypotheses of the PAP.