# Government Audits of Municipal Corruption and Belief Updating: Experimental Evidence

Felipe Torres-Raposo[*]

London School of Economics and Political Science

Raymond Duch

Nuffield College

November 28, 2025

**Abstract**

Government audits are widely used to expose public malfeasance, but it is unclear whether such information changes citizens' beliefs or political behaviors. We conducted a field experiment in Chile in which over 5,000 citizens were randomly assigned to view short WhatsApp video summaries of recent municipal audit findings. The audit information caused citizens to significantly update their beliefs about local government corruption, with similar effects across partisan groups. However, these belief changes did not vary meaningfully with the magnitude of malfeasance reported in the audits. Moreover, the information had no detectable effect on related behaviors such as voting or municipal donations. Our findings suggest that while audit disclosures shift public beliefs, they may not spur corresponding behavioral change.

**Keywords:** Corruption, Belief Updating, Audits

[*] f.torres-raposo@lse.ac.uk

Most citizens readily express opposition to corruption (de Sousa et al., 2023). Yet scholars continue to grapple with understanding how the average citizen forms beliefs about "corruption." This is challenging partly because government corruption is so diverse: It can, for example, be active versus passive (Bandiera et al., 2009); it can be severe or petty (Bauhr & Charron, 2018; Brollo et al., 2013; Martinelli, 2022); and it can range from bribery (Erlich et al., 2025) to criminal activity (Buckley et al., 2022). Government audit reports may capture some of this diversity but typically they provide an overall quantitative measure of corruption that attempts to summarize this diversity. Our field experiment demonstrates that this quantitative information generated by audit reports has a causal effect on the corruption beliefs of average citizens.

Audit reports serve as a credible source of information regarding the level of malfeasance detected within government institutions. Independent audits of government activities are widely endorsed; many countries have adopted them (OECD, 2016); and aid funding agencies support their establishment (Berliner & Wehner, 2022). Recent studies established that audits are an accurate measure of, as well as an effective strategy for mitigating, corruption (Denly, 2022; Lagunes, 2021).

In partnership with the NGO Chile Transparente and the Comptroller General Office (CGO), we assigned video treatments, via WhatsApp, summarizing municipality audit results released by the CGO just prior to the experiment. Treated subjects observed a video reporting 1) the level of malfeasance in pesos; 2) the cost of malfeasance in terms of foregone flu vaccines; and 3) whether levels have improved or deteriorated since the previous audit. The placebo group saw a video unrelated to corruption or malfeasance. The effects of this audit information on corruption beliefs is estimated one week and one month post-treatment.

We find that participants' initial beliefs about municipal corruption (their priors) were strong predictors of their later beliefs (posteriors) after the intervention, indicating that citizens heavily anchor on preexisting impressions. However, those priors were not particularly accurate, leaving ample room for information to shift perceptions. A short, mobile-delivered video summarizing local audit results significantly increased respondents' belief that malfeasance was occurring in their municipality. In other words, treated participants revised their corruption estimates upward on average, even as participants' beliefs, again on average, showed some natural regression toward the mean. The information treatment thus moved corruption perceptions above and beyond any baseline tendency to adjust beliefs toward the average. Importantly, these effects did not differ by partisanship: co-partisans and non-co-partisans of the incumbent mayor were equally influenced by the corruption video. The treatment's impact was also robust over time – treated subjects re-surveyed a month later continued to report higher perceived corruption than the control group. We find no evidence though that this belief updating is correlated with costly donations to the local municipality nor with voting for the incumbent. Overall, the intervention clearly shifted public beliefs about local malfeasance, even if citizens did not absorb the precise numeric details of the audit statistics. A brief, accessible information treatment was enough to alter corruption perceptions in a lasting way.

This study contributes to several strands of research by indicating how, and when, citizens update corruption beliefs. First, it speaks to the literature on political accountability and corruption information that reports mixed effects of malfeasance information on voter attitudes and behavior. Some studies find that voters respond to malfeasance information by punishing corrupt politicians at the polls (Ferraz & Finan, 2008); others observe demobilization or cynicism in response to anti-corruption messages (Cheeseman & Peiffer, 2021; Chong et al., 2015); while some note that information effects are conditioned by partisan or ethnic bias (Adida et al., 2020).

By specifically focusing on belief formation, our experiment sheds light on these disparate findings. We show that average citizens incorporate features of the audit information into their beliefs, but in an incomplete manner. Second, our results inform broader debates on belief updating and motivated reasoning in political contexts. Classic theories warn that partisans will resist or discount unwelcome information (Kunda, 1990; Taber & Lodge, 2006), yet in our study co-partisans and non-co-partisans updated their beliefs in the same direction. This uniform treatment effect aligns with recent work suggesting persuasion is often symmetric across partisan groups (Coppock, 2023) and is consistent with models of rational updating in political economy (DellaVigna & Gentzkow, 2010). Credible, non-partisan information about corruption can change minds even in a polarized setting.

Finally, we contribute to a growing body of research on how the public responds to quantitative information interventions. Recent experiments in economics show that providing factual data can shift people's beliefs about macro-economic conditions or policies, though often with only partial uptake of the numeric content (Coibion et al., 2023; D'Acunto & Weber, 2024; Roth & Wohlfart, 2020) . Our findings parallel this pattern: citizens updated their corruption estimates in the expected direction when presented with audit statistics, but they did not fully internalize the quantitative details. In sum, we offer new evidence on how ordinary citizens form posterior beliefs when confronted with numeric evidence of political corruption.

Our paper proceeds as follows. We first present the experimental design and then discuss how we estimate and model belief updating. The third section presents the results of the experiment. Finally, we suggest directions for future studies of the public's beliefs about corruption.

# Corruption Belief Updating

Our pre-registered conjecture is that the irregularities reported in the Contraloria audits cause updating of corruption beliefs. We review our strategy for measuring audit irregularities and corruption beliefs. Finally, we discuss models of belief updating that inform our experimental design and analysis of the results.

**Corruption in Chile.** According to Transparency International's Corruption Perceptions Index (Transparency International, 2023), Chile ranked 27th out of 180 countries. The index reflects perceptions of public sector corruption, drawing on 12 data sources that assess rule of law, institutional strength, and financial integrity. In spite of these relatively high scores, Chile experienced a series of high-profile political corruption scandals in the years preceding our study. These included illicit campaign financing involving a substantial number of parliamentarians and influence-peddling by family members of the sitting president. Corruption was a salient political issue in Chile. LAPOP's

AmericasBarometer surveys (LAPOP, 2021), shown in Figure A.2 in the Appendix, rank corruption as the second most critical national issue in 2017, dropping to fourth in 2019, and falling to tenth in 2021, surpassed by growing concerns related to the COVID-19 pandemic, such as health and unemployment.

**Audit Reports.** Audit reports of governmental malfeasance are hypothesized to affect the public's beliefs about corruption. A case in point is the Brazilian Controladoria-Geral da União that reports the total amount of federal funds audited, along with an itemized list describing each irregularity such as the diversion of public funds to irregular and illegal procurement practices. Ferraz and Finan (2008) find that the probability of re-election declines from 53% when no irregularities are detected to 20% when three irregularities are uncovered. They argue that this decline in electoral support operates through voters' updating of their beliefs regarding the incumbent's involvement in corruption.

Similarly, the Office of the Comptroller of Puerto Rico identifies malfeasance in diverse from: from fraud in procurement, fake receipts, illegal hiring of public officials, and over-invoicing. Bobonis et al. (2016) show that the *timely* release of these audit results in Puerto Rico, particularly when they occur close to elections, has significant short-term effects in curbing both corruption and incumbents' re-election rates. In municipalities where corruption was uncovered, re-election rates fell from 40% to 22%. These effects are mediated by the extent to which voters learned about the audits' findings through primarily conventional media channels. Neither the Bobonis et al. nor the Ferraz and Finan study provide direct evidence that malfeasance information results in the updating of corruption beliefs.

The Federal Auditor's Office (ASF) in Mexico reports the share of spending of federal funds diverted into unauthorized projects, with a focus on the spending, accounting, and management of these resources. Using data from the ASF Arias et al. (2022a) demonstrate that voters update their beliefs when presented with two pieces of information: the total amount of federal transfers received by each municipality and the proportion of those transfers that were used for unauthorized purposes. Similarly, Chong et al. (2015) employed the same audit data and analogous informational treatments, as in Arias et al. (2022b), to evaluate the effects of audit-based information on political attitudes and behavior. Their findings suggest that while malfeasance information did not significantly alter citizens' perceptions of corruption or support for the incumbent, it did lead to lower electoral turnout.[1]

The Chile audit agency, the CGO, has a reputation among citizens and civil servants as being non-partisan and reliable.[2] Chile's CGO audit agency scores between 8.0 and 8.5 on the *Supreme Audit Institutions Independence Index* constructed by the World Bank (2021) ranking it in the top 65% of countries with audit agencies. It has extensive access to records and information on most governmental activities at the national and local levels. The audit agency has a constitutional mandate to report, disseminate, and act on the results of these audits. As Denly (2022) shows, audit agencies with similar scores to Chile's CGO, such as the Indian and Guatemalan audit agencies, are endowed with considerable independence to audit all public funds. [3]

Over the last decade, the CGO has conducted annual audits of approximately 180 of the 345 Chilean municipalities. Audits take around 6 months. Each annual audit typically focuses on different activities, such as procurement, hiring processes, or municipal finances. Audited municipalities are selected based on a scoring system incorporating factors such as budget size, transfers to the private sector, and results from previous audits. The factors and weights incorporated into the final scoring change from one audit plan to another.

The CGO conducted their municipal audits from January through December 2020. The audit reports were released months before the May 2021 municipal election. As the CGO released audit results for the 116 municipalities, we produced a customized 50-second video for each mu-

[1] Jablonski et al. (2021) also compiled information from audits carried out by enumerators on primary schools, roads, health care, and water access quality in Ugandan villages. The author created indices of "service quality" for each village and informed voters whether their village performed better or worse than other villages within the same district. The authors find no effects of their treatments on changing citizens' beliefs about the incumbent's integrity and effort, as well as voting outcomes.

[2] Based on annual representative surveys by the Council for Transparency (2018–2020), the CGO consistently ranked among the most trusted government institutions (para la Transparencia, 2018, 2019, 2020).

[3] In Section A.6, in the Appendix, we provide a more extensive discussion of the audits conducted by the CGO.

nicipality that summarized the reported irregularities.[4] For each municipality, we also produced a placebo video that provides neutral statistics on the municipality, such as population and size of the commune.

**Corruption Beliefs.** Our outcome variable is corruption beliefs. Most measures of corruption beliefs employ a qualitative indicator. These measurements can include individual perceptions of levels or changes in the amount of "corruption" (Peiffer, 2020; Reinikka & Svensson, 2006), bribery (Corbacho et al., 2016; Letki et al., 2023), fraud (Zhou & Oostendorp, 2014), nepotism (Gagliarducci & Manacorda, 2020), waste (Bandiera et al., 2009), or vote-buying (Gonzalez-Ocantos et al., 2014; Hicken et al., 2015). We adopt a qualitative measure consistent with the measurement approach used by (Arias et al., 2022a; Enríquez et al., 2024).

Audit reports are typically intended to provide a quantitative assessment of government malfeasance. Thus, our two primary belief measures are designed to capture the quantitative dimension of these reports. First, we elicit respondents' perceptions of local government malfeasance using a distributional question that assigns probabilities to pre-specified bins of malfeasance levels, following the approach of Arias et al. (2022b) and Arias et al. (2019). Second, we include a belief measure that mirrors those used in the inflation expectations (Roth & Wohlfart, 2020) and wage expectations literature (Jäger et al., 2024), where respondents are asked to indicate their best guess from a wide set of discrete ordinal options.

**Corruption Belief Updating.** The field experiment has a treatment arm designed to estimate the causal effect of malfeasance information on corruption belief updating. There are also features of the information treatment that are not randomly assigned to subjects but provide insights into corruption belief updating. We briefly highlight our conjectures as to how features of the information treatment shape corruption belief updating.

*Information Treatment.* The randomly assigned information treatment in our experiment is a

[4] The videos for each of the 116 municipalities and the data used in these videos are available at: https: (link omitted to preserve anonymity)

quantitative narrative about local governmental corruption. It signals that public officials commit malfeasance. Our pre-registered hypothesis is that subjects will update their corruption beliefs when they receive audit information about local municipal government malfeasance. Being informed of municipal government malfeasance is hypothesized to result in higher beliefs about corruption.

These narratives provide details about malfeasance that are customized to reflect their municipality's audit report. Respondents learn about magnitudes of malfeasance and whether performance has improved or gotten worse. Hence we both signal the presence of malfeasance in local government but also quantitative details that can inform subjects' belief updating about local corruption. But it is challenging to effectively communicate nuances regarding corruption to average citizens. We observe this with anti-corruption messaging designed to change corruption attitudes and corrupt behavior; these efforts can either have a null effect or in some circumstances exactly the opposite intended outcome (Cheeseman & Peiffer, 2021; Chesseman & Peiffer, 2024).

While subjects are randomly assigned to receive an audit report versus a placebo video treatment, the specific quantitative details regarding municipal government malfeasance are not randomly assigned. While not well-identified, the response of subjects to these quantitative details provides suggestive insights into belief updating. There are features of the quantitative information treatment that we expect to correlate with the direction and magnitude of belief updating and that speak to a number of conjectures regarding corruption beliefs.

*Benchmarking.* The information treatment provides respondents with a comparison of their municipality's current audit performance with previous audit results. This notion that citizens engage in benchmarking is a foundational assumption of many models of responsibility attribution (Barro, 1973; Ferejohn, 1986; Fiorina, 1981). There is evidence of benchmarking incumbent government performance; certainly concerning economic performance (Arel-Bundock et al., 2021; Aytaç, 2018; Kayser & Peress, 2012); but also in other policy areas such health (Becher et al., 2023; Duch et al., 2025; Rodríguez et al., 2025) and education (Charbonneau & Van Ryzin, 2015). Studies though suggest that benchmarking is fraught with difficulties (Arel-Bundock et al., 2021;

Healy & Lenz, 2014).

A challenge, particularly in the study of corruption beliefs, is determining the referents of this benchmarking. Spatial benchmarking, for example, juxtaposes incumbent malfeasance information with comparisons from other similar jurisdictions (Banerjee et al., 2014; Bhandari et al., 2019; Enríquez et al., 2024). Temporal benchmarking, on the other hand, compares incumbent performance with their historical performance or with previous outcomes under a challenger (Arias et al., 2025; Avenburg, 2019; Botero et al., 2015; Breitenstein, 2019). The expectation is that these benchmarks inform corruption beliefs by providing a reference point in assessing levels of malfeasance but the evidence of their effectiveness is mixed.

*Magnitudes.* The ability of individuals to make sense of malfeasance metrics is an important micro-foundation for empirical studies of corruption (Arias et al., 2022a; Boas et al., 2019; Chong et al., 2015; Enríquez et al., 2024; Ferraz & Finan, 2011; Larreguy et al., 2020). A concern though is that average citizens struggle to correctly assess these quantitative policy outcome metrics. We see this, for example, with respect to crime statistics (Esberg & Mummolo, 2018; Larsen & Olsen, 2020); immigration (Blinder & Schaffner, 2019; Hopkins et al., 2019), government debt (Roth et al., 2022), and trade balances (Mutz, 2021; Rodríguez Chatruc et al., 2021).

We are able to assess whether, and how, reported magnitude of malfeasance is incorporated into belief updating. As many studies suggest, the framing of these metrics affects engagement and comprehension; potentially calling into question whether subjects are actually “treated”. This is the case for quantitative indicators of malfeasance because there are many possibilities for framing the metrics: The Enríquez et al. (2024) information treatments, for example, consisted of the percentage of audited funds subject to irregularities. Similarly, Boas et al. (2019) informed voters of the rate of funds rejected by the local audit agency. A challenge for scholars in this field, including us, is settling on a framing and presentation of this metric that is engaging and comprehensible for the average citizen. We designed a measurement strategy for malfeasance magnitude that we believe optimizes on engagement and comprehension.

We conjecture that corruption beliefs will update based on the magnitude of municipality government malfeasance. Respondents in the trial will observe magnitudes of reported malfeasance that will vary significantly depending on their municipalities. Our expectation is that the magnitude of belief updating will be correlated with the amounts of reported malfeasance in the subject's municipality.

*Information Gap.* Engaging with and responding to the magnitude of reported malfeasance is a necessary condition for acquiring accurate beliefs about corruption. These cognitive steps help reduce the gap between subjective beliefs and objective reality (Cavallo et al., 2017; D'Acunto & Weber, 2024). There is some evidence that, under certain conditions, average citizens receive ground-truth information and subsequently update their beliefs in a direction that narrows this gap (Jäger et al., 2024; Tappin et al., 2020). However, other studies offer less encouraging findings: corrective information can sometimes backfire, exacerbating the discrepancy between beliefs and reality (Carey et al., 2025; Nyhan & Reifler, 2010; Zappalà, 2024). Our interest is in whether credible information regarding malfeasance causes citizens to form more accurate beliefs about corruption levels (Liang, 2025; Sanna & Lagnado, 2025).

The experimental design allows us to calibrate this gap between beliefs and ground truth. Ground truth is derived from audit agency data and we elicit beliefs about this ground truth with incentived survey questions. We measure these incentivized corruption beliefs, for control and treated subjects, both pre-treatment (measured four months before the intervention) and post-treatment. Our conjecture is that providing subjects with audit agency data about their municipality will reduce this gap between beliefs and ground truth. Subjects have a financial incentive to use this information to narrow this information gap. But even given the financial incentives a host of factors may compromise informed updating.[5]

*High/Low Malfeasance Priors.* Belief updating may differ for individuals with extreme priors – either high or low (Fan et al., 2024). This might result from anchoring and conservative updating at the extremes (Coutts, 2019; Eil & Rao, 2011; Kovach, 2021) or expectations reference points (Coibion & Gorodnichenko, 2015). Several corruption studies have provided evidence that belief

[5] Both quantitative outcomes were incentivized: respondents could earn a 20% bonus on their survey participation payment for answering these questions correctly.

updating is conditional on citizens' priors relative to what is reported in the signal. A case in point is (Arias et al., 2022a), who report that citizens' responsiveness to information from corruption audits in Mexico is conditional on their priors' beliefs about malfeasance. The authors find that voters who initially believed corruption was high in their constituency revised their beliefs more favorably after receiving information about the incumbent's level of malfeasance, interpreting the evidence as indicating less corruption than expected. In contrast, voters who initially believed there was little corruption became more pessimistic, viewing the same information as evidence of greater malfeasance than they had assumed.

A contributing factor may be diminishing returns to information for people with very high or low corruption priors. Individuals with relatively negative priors regarding malfeasance in their local government respond to negative versus positive corruption signals differently than those with relatively positive priors. The effect of providing negative information about the malfeasance of a municipal government will be decreasing in individuals' prior beliefs that the government is malfeasant. Similarly, the effect of providing positive information would be decreasing in the individuals' prior beliefs that the government is not malfeasant. Those with negative (positive) malfeasance priors may already anticipate audit results that report negative (positive) information about irregularities in their municipal government, and hence, there is minimal updating reflected in their posteriors.

We observe subjects with a wide range of corruption priors and conjecture that those at the extremes of the distribution will exhibit patterns of belief updating that are less likely to shrink gaps between beliefs and ground truth.

*Partisanship.* Partisan motivated reasoning may color how individuals process policy-related information (Kunda, 1990; Mian et al., 2021; Peterson & Iyengar, 2021). Thaler (2024) presents experimental results suggesting that partisanship conditions Bayesian updating with respect to a range of quantitative policy metrics. Recent research also suggests that partisanship affects how individuals respond politically to information about government corruption (de Figueiredo et al., 2022; Elia & Schwindt-Bayer, 2022; Solaz et al., 2019). This implies that partisanship conditions belief updating when individuals receive new quantitative information about malfeasance by their local government (Anduiza et al., 2013; Cornejo, 2022; Cubel et al., 2024).

An alternative conjecture is that while there are partisan differences in beliefs (i.e., priors) about levels of corruption in their municipality, both co-partisans and non co-partisans will update their beliefs similarly in response to factual information about corruption. This "parallel persuasion" argument Coppock (2023) suggests, first, that co-partisans of the governing municipal party will perceive the levels of municipal government malfeasance to be much lower than the levels perceived by non co-partisans. Second, it implies that our malfeasance treatment will have similar effects on co-partisans and non co-partisans.

*Behavior.* Ultimately, we are interested in corruption belief updating because these changing beliefs affect behavior; in particular, political behavior. Again, while these updated beliefs are not randomly assigned their correlation with behaviors can be informative. We will focus on two political behaviors: public goods contributions and incumbent vote. Government malfeasance is likely correlated with the public's enthusiasm for contributing to public goods provision (Beekman et al., 2014; Campos-Vazquez & Mejia, 2016; Jahnke & Weisser, 2019; Peyton, 2020). There is also an extensive literature exploring the link between governmental malfeasance and vote for incumbent candidates (Arias et al., 2022a; Banerjee et al., 2014; Boas et al., 2019; Buntaine et al., 2018; Chong et al., 2015; Dunning, Grossman, Humphreys, Hyde, et al., 2019; Enríquez et al., 2024; Ferraz & Finan, 2008). Our expectation is that the belief updating we observe would be correlated with incumbent vote preferences.

# Experimental Design

**Recruitment.** In anticipation of the 2020 annual municipal audits, we recruited a subject pool of approximately 49,883 subjects primarily via Facebook, Instagram, Twitter, and standard recruitment approaches for panel surveys such as posters in public spaces, person recruitment efforts, and advertisements in newspapers or other local media outlets.[6] In January 2019, the CGO selected

[6] Full details on the recruitment campaign (including ads) are described in Section A2 in the Appendix.

116 municipalities for the 2020 audit program. Our pool contained 46,723 residents in the 116 municipalities. Table A.1 in the Appendix benchmarks the demographics of the final sample against the Chile census and the National Social-Economical Characterization Survey statistics. Our final sample has an age profile similar to the census; is more female and is significantly better educated. In January-March 2021, we then conducted a pre-treatment survey with 5,528 participants from our pool who were residents in these 116 municipalities. [7]

**The Malfeasance Signal.** Treated subjects are informed of malfeasance audit outcomes with a short 50-second video delivered via WhatsApp. 85% of respondents in our sample reported using WhatsApp multiple times a day, and 75% claimed to receive political information through this medium, which is considerably higher relative to other social media platforms such as Facebook (44%), Instagram (49%) and TikTok (22%).

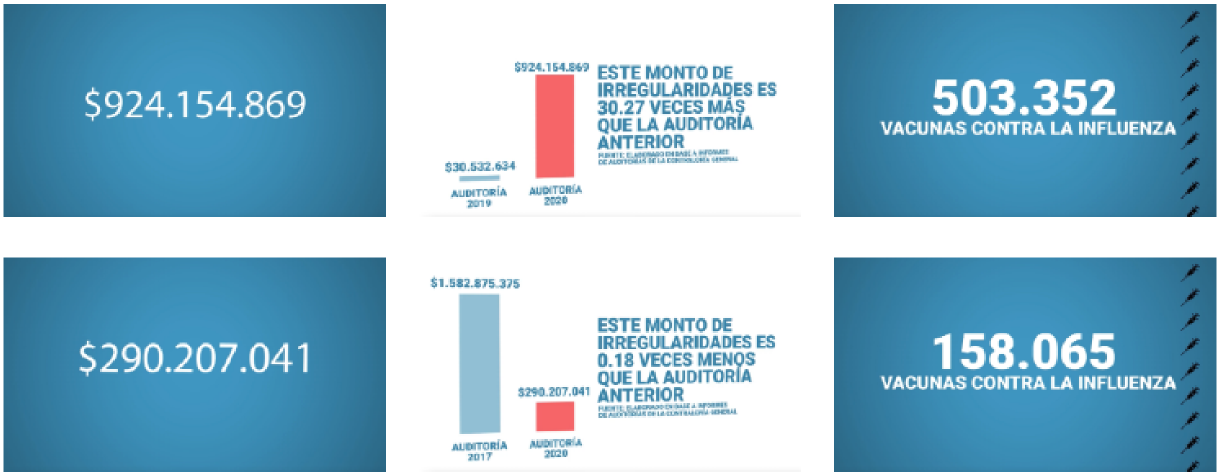
We designed an experiment to identify the messaging that effectively signaled government malfeasance. As pointed out before, there are diverse strategies for designing corruption information treatments. For example, the magnitude of malfeasance has been expressed in overall levels, in terms of percentages of the budget (Buntaine et al., 2018) or as foregone expenditures (Arias et al., 2022a; Larreguy et al., 2020). Bhandari et al. (2019), for example, find that "temporal" benchmarking (comparing current malfeasance in a district performance with the district's previous levels) is more likely to affect updating than is the case with "spatial" benchmarking (comparing a district's malfeasance levels with those of other districts). Prior to producing the information treatment videos, we implemented a Chatbot experiment (n = 3996) on WhatsApp that evaluated 24 different messaging strategies. Subjects were randomly assigned six different video messages and asked to evaluate each of them. Results suggested that messages identifying "foregone resources" and incorporating a "temporal" benchmark were the most effective framing of malfeasance information (Citation omitted to preserve anonymity). We implemented these information treatment frames in

[7] Subjects in the pool were given a window of more than 4 weeks to reply to the invitation to participate. Initially, 6,050 eligible respondents answered the baseline survey. A total of 5,528 respondents participated in the final study – we excluded 10 municipalities because the audit reports were not released on time by the CGO or were published very close to the election date. Hence, the effective response rate was about 25 percent.

the full-scale field experiment.

Figure 1 presents the three frames of the treatment video. The videos are customized to reflect the CGO audit results for the participant's municipal government. First, respondents received information about the total amount of malfeasance found in the 2020 audit report. A second frame compares current reported malfeasance levels with those reported in the municipality's previous audit. The third frame expresses malfeasance magnitude as the number of foregone flu vaccines for the municipal population. Figure 1 illustrates both a worsening (top row) and improving (bottom row) audit result. [8]

Figure 1: Screenshots Depicting the Video Treatment.


$924.154.869
$924.154.869
ESTE MONTO DE IRREGULARIDADES ES 30.27 VECES MÁS QUE LA AUDITORÍA ANTERIOR
$30.532.634
AUDITORÍA 2019
AUDITORÍA 2020
503.352
VACUNAS CONTRA LA INFLUENZA
$290.207.041
$1.582.875.375
ESTE MONTO DE IRREGULARIDADES ES 0.18 VECES MENOS QUE LA AUDITORÍA ANTERIOR
$290.207.041
AUDITORÍA 2017
AUDITORÍA 2020
158.065
VACUNAS CONTRA LA INFLUENZA

*Note:* These are screenshots of the video information treatments that subjects received. The top row has screenshots of respondents in municipalities with rising malfeasance. The first part of the video reports the amount of malfeasance found in their municipality. The second screenshot is a temporal comparison – corruption has increased over time in this municipality. The third screenshot on the top row shows "foregone losses" expressed as the number of influenza vaccines. The three images on the bottom row show the same information, although for a municipality with decreasing levels of irregularities over time (we label this "positive" information).

**Treatment Assignment.** Figure 2 provides a summary of the experiment design. Treatment assignment was conducted within each of the 116 municipalities. Our initial pilot results indicated that respondents' prior beliefs are strongly correlated with their posteriors. Subject's malfeasance priors (both qualitative and quantitative) were measured in the pre-treatment survey. We constructed *High* and *Low* blocks using these subjective malfeasance scales (from 1 to 10). Within

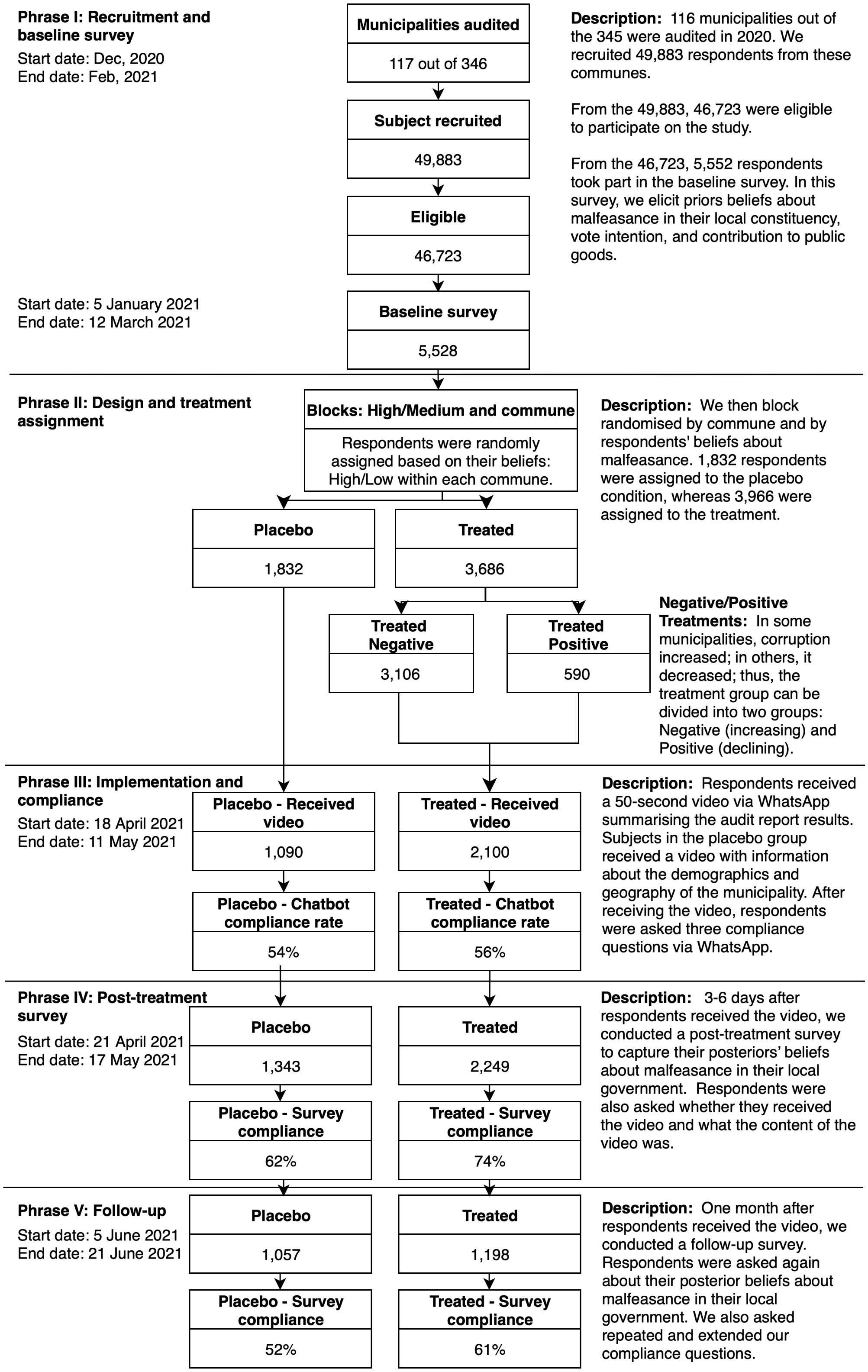
[8] Screenshots of the placebo video can be found in Appendix Figure A.1.

each municipality, we blocked on high versus low malfeasance priors; subjects within each block were randomly assigned to either their municipal malfeasance treatment video or the placebo condition video. We blocked on priors beliefs, as based on our pilot results, subjects' priors were strongly correlated with posteriors. This design element allowed us to create homogeneous sets of respondents based on their priors and improve the precision of our estimates (Gerber & Green, 2012).

This resulted in 3,696 participants assigned to treatment and 1,832 assigned to the placebo condition. The oversized treatment group is an artifact of an initial design that anticipated two treatment arms plus a placebo arm. The final implementation consisted of only one treatment arm and the placebo without adjusting the sample sizes of the final treatment and placebo arms. The treatment and placebo videos were sent to these 5,528 participants via WhatsApp.[9] A post-treatment survey was administered to participants 3-6 days after they were treated with the videos. 3,592 subjects responded to the survey: 2,249 out of 3,696 treated (61%) and 1,343 out of the 1,832 placebo subjects (73%). Finally, a follow-up survey was administered one month after the video treatments, with 1,198 out of 3,696 treated (32%) responding and 1,057 out of the 1,832 placebo subjects (58%) responding.

[9] Immediately after respondents watched the video, they were asked a series of compliance and attention check questions. Out of the 1,832 respondents in the placebo group, 1,055 answered correctly the follow-up Chatbot questionnaire (57%). Of the 3,696 respondents in treatment, 1,961 answered correctly the Chatbot attention check questions (53%).

Figure 2: Flow Chart of Experimental Design.


Phrase I: Recruitment and baseline survey
Start date: Dec, 2020
End date: Feb, 2021
Municipalities audited
117 out of 346
Subject recruited
49,883
Eligible
46,723
Start date: 5 January 2021
End date: 12 March 2021
Baseline survey
5,528
Phrase II: Design and treatment assignment
Blocks: High/Medium and commune
Respondents were randomly assigned based on their beliefs: High/Low within each commune.
Placebo
1,832
Treated
3,686
Treated Negative
3,106
Treated Positive
590
Phrase III: Implementation and compliance
Start date: 18 April 2021
End date: 11 May 2021
Placebo - Received video
1,090
Treated - Received video
2,100
Placebo - Chatbot compliance rate
54%
Treated - Chatbot compliance rate
56%
Phrase IV: Post-treatment survey
Start date: 21 April 2021
End date: 17 May 2021
Placebo
1,343
Treated
2,249
Placebo - Survey compliance
62%
Treated - Survey compliance
74%
Phrase V: Follow-up
Start date: 5 June 2021
End date: 21 June 2021
Placebo
1,057
Treated
1,198
Placebo - Survey compliance
52%
Treated - Survey compliance
61%

**Description:** 116 municipalities out of the 345 were audited in 2020. We recruited 49,883 respondents from these communes.

From the 49,883, 46,723 were eligible to participate on the study.

From the 46,723, 5,552 respondents took part in the baseline survey. In this survey, we elicit priors beliefs about malfeasance in their local constituency, vote intention, and contribution to public goods.

**Description:** We then block randomised by commune and by respondents' beliefs about malfeasance. 1,832 respondents were assigned to the placebo condition, whereas 3,966 were assigned to the treatment.

**Negative/Positive Treatments:** In some municipalities, corruption increased; in others, it decreased; thus, the treatment group can be divided into two groups: Negative (increasing) and Positive (declining).

**Description:** Respondents received a 50-second video via WhatsApp summarising the audit report results. Subjects in the placebo group received a video with information about the demographics and geography of the municipality. After receiving the video, respondents were asked three compliance questions via WhatsApp.

**Description:** 3-6 days after respondents received the video, we conducted a post-treatment survey to capture their posteriors' beliefs about malfeasance in their local government. Respondents were also asked whether they received the video and what the content of the video was.

**Description:** One month after respondents received the video, we conducted a follow-up survey. Respondents were asked again about their posterior beliefs about malfeasance in their local government. We also asked repeated and extended our compliance questions.

**Outcomes.** Table 1 presents the treatment and placebo group mean values and standard errors for our three outcome variables measured at each survey wave.[10] The Qualitative outcome measurement question is adapted from Enríquez et al. (2024): “How much of the municipal government expenditures are corrupt?” (10-point scale from none to all). It’s average value for treated subjects rises from 4.49 in pre-treatment to 5.34 post-treatment and then declines somewhat to 5.15.

Building on recent findings regarding incentives and accuracy of beliefs (Rathje et al., 2023; Zimmermann, 2020), we incentivized the two quantitative outcomes: *Distribution* and *Resources*. The *Distribution* variable is subjects’ reported ranking of their municipality in terms of levels of malfeasance: Low (1); Moderate (2); High (3); and Very High (4). They are presented with the actual distribution of Chilean municipalities in four categories. Participants earn 300 CLP (Chilean Pesos) if they correctly categorize their municipality.[11] For treated subjects, in pre-treatment, the average Distribution score is 1.91, just below moderate malfeasance, rising to 2.29 in the initial post-treatment survey and then declining to 2.19 in the follow-up.

For the incentivized *Resource* measure, we ask participants to indicate how much out of every 1000 CLP municipality expenditures is associated with corrupt activities. Participants earn an additional 300 CLP if they get the correct answer. This Resource measure has been converted to ordinal scores ranging from 1 (”Less than 100 CLP”) to 2 (”Between 100 and 200 CLP) until 10 (”More than 900 out of 1000 CLP”). For treated subjects, in pre-treatment, the average Resource score is 3.85; in post-treatment, the average increases to 4.57, but it declines in the follow-up to 4.22.

Table 1 presents the treatment and placebo group mean values and standard errors for the outcome variables measured at each of the three survey waves. For all three outcome variables, we observe that average values rise in the post-treatment wave for treated subjects and then decline somewhat in the third follow-up wave. The average values of subjects in the placebo condition are similar across all three waves.

[10] Several recent studies have highlighted the importance of robust measurement strategies to avoid experimental measurement error (Gillen et al., 2019; Haaland et al., 2023).

[11] This incentive represented about 20% increase in their subject payment, if they answered correctly both incentivized questions.

Table 1: Descriptive Mean Values for Malfeasance Belief Outcome Measures and Covariates.

| | Survey Wave | | | | | |
|---|---|---|---|---|---|---|
| | Pre-Treat | | Post-Treat | | Follow-up | |
| | Placebo | Treatment | Placebo | Treatment | Placebo | Treatment |
| **Qualitative Measures** | | | | | | |
| Subjective malfeasance scale (1-10) | 4.47 | 4.49 | 4.60 | 5.34 | 4.39 | 5.15 |
| | (2.44) | (2.44) | (2.35) | (2.34) | (2.30) | (2.29) |
| Certainty subjective malfeasance scale (1-10) | 5.44 | 5.39 | 5.28 | 5.74 | 5.46 | 5.52 |
| | (2.72) | (2.70) | (2.72) | (2.59) | (2.72) | (2.59) |
| **Quantitative Measures - Incentivized** | | | | | | |
| Resources (1-10) | 3.83 | 3.85 | 3.78 | 4.57 | 3.52 | 4.22 |
| | (2.36) | (2.39) | (2.36) | (2.60) | (2.22) | (2.48) |
| Distribution (1-4) | 1.90 | 1.91 | 1.93 | 2.29 | 1.89 | 2.19 |
| | (0.78) | (0.79) | (0.79) | (0.87) | (0.78) | (0.84) |
| **Behavioral Measures** | | | | | | |
| 500 CLP Municipal Donation | 32% | 33% | 17% | 15% | 11% | 10% |
| | (1.09%) | (0.77%) | (1.27%) | (0.78%) | (1.42%) | (1.34%) |
| **Covariates** | | | | | | |
| Female | 64.79% | 64.69% | 64.78% | 64.61% | 64.81% | 66.61% |
| | (1.11%) | (0.78%) | (1.3%) | (1%) | (1.46%) | (1.35%) |
| Education (Yrs) | 15.09 | 15.05 | 15.10 | 15.04 | 15.06 | 15.03 |
| | (0.06) | (0.04) | (0.08) | (0.06) | (0.08) | (0.08) |
| Age | 36.42 | 35.61 | 36.41 | 35.37 | 36.45 | 35.17 |
| | (0.31) | (0.21) | (0.37) | (0.28) | (0.43) | (0.38) |
| Partisanship (1 to 10) | 4.55 | 4.43 | 4.58 | 4.38 | 4.58 | 4.31 |
| | (0.05) | (0.03) | (0.06) | (0.04) | (0.06) | (0.06) |
| Co-partisanship | 28% | 27% | 28% | 26% | 28% | 26% |
| | (1.05%) | (0.73%) | (1.22%) | (0.93%) | (1.38%) | (1.27%) |
| Income (USD) | 678 | 648 | 656 | 621 | 604 | 556 |
| | (30) | (18) | (28) | (24) | (29) | (21) |
| Sample | 1832 | 3696 | 1343 | 2249 | 1057 | 1198 |

*Note:* This table reports mean values for all outcomes and key covariates across waves and by treatment group. For *Behavioral* outcomes, percentages indicate the share donated to respondents' municipalities. *Female* reflects the proportion of women; *Education* is measured in years. *Partisanship* ranges from 0 (”left”) to 10 (”right”). *Co-partisanship* is the share intending to vote for the incumbent mayor. *Income* is based on self-reported brackets, using bracket midpoints to estimate individual income. Standard errors are in parentheses.

Respondents are asked to donate 500 CLP (about 15 percent of their earnings) to the provision of a local public good. In the initial wave, about one-third of treated respondents donated 500 CLP – this drops to 15 percent in the post-treatment survey for treated subjects and then drops to 10 percent in the one-month follow-up.

Table 1 summarizes covariate values. Our sample is disproportionately female. Most of our sample had completed primary and secondary education and some form of tertiary education. The mean partisanship across all waves is Left-leaning with a score of 4.4 measured on a Left-Right scale ranging from 1 to 10. Approximately one-quarter of the respondents indicated they would vote for the incumbent mayor (*Co-partisan*). Average individual monthly income is in the range of $648-$678 in pre-treatment and declines somewhat over the subsequent waves.

**Balance and Attrition.** Respondents are observed at three different waves of the trial: baseline when treatment is assigned and implemented; one-week post-treatment; and one-month post-treatment. *Balance* on covariates is assessed in Appendix Figure A.3 that compares standardized mean differences for the treatment and placebo arms for each of the three waves of the trial. We employ indicative balance tolerance levels of 0.1 (Stuart et al., 2013). With only a few exceptions, the unadjusted sample standardized mean differences for covariates across the three comparisons fall within this 0.1 threshold. Table A.2 in the Appendix reports the estimates from regressing a dummy treatment assignment variable on pre-treatment covariates, confirming balance across treatment arms.

The Appendix examines on how differential *attrition* between treatment at baseline and post-treatments (one week and one month) affects the composition of the two treatment arms. Table A.3 in the Appendix presents the results of logit regressions of compliance in post-treatment waves (subjects have a value of 0 if they answered and 1 if they did not) on the treatment arm dummy variable. Across all models, the treatment arm coefficients are precisely estimated and positive, suggesting higher attrition rates in the treatment arm. In Table A.3, which includes a complete set of covariate and covariate-treatment interaction terms, we observe that attrition positively correlates with income, higher education attainment, and civic knowledge.

In Appendix Table A.4, we compare baseline priors measures for individuals who attrited from the study with those who completed both the post-treatment and follow-up surveys. We find that attriters report higher baseline perceptions of malfeasance for the Qualitative outcome, though we detect no significant differences in quantitative outcomes. In terms of demographic characteristics, the two groups are broadly similar in terms of gender, education, age, and partisanship. However, we observe notable differences in income levels.

Despite this pattern of differential attrition between treatment and placebo groups, the overall covariate balance remains intact, as shown in Figure A.3. We note that the study's overall attrition rate remains within the upper range observed in comparable field experiments, as documented by Ghanem et al. (2023), with a differential attrition rate of 13%. To formally account for potential biases arising from attrition, we estimate models predicting attrition status and compute inverse probability weights (IPW), as detailed in the Appendix. Table A.7 presents the treatment effect estimates using these weights alongside the unweighted results from Table 2. The treatment effects are consistent across both approaches, providing further reassurance that attrition does not meaningfully distort the main findings. We provide a more extensive discussion of differential attrition in Section A.3.

A post-treatment survey was administered to participants 3-6 days after they were treated with the videos. We asked these subjects whether they recalled viewing the treatment videos: 846 (63% of those responding to the post-treatment survey) indicated they had viewed the Placebo video, and 1,670 (74% of those responding to the post-treatment survey) indicated viewing the treatment videos. In Table A.8 in the Appendix, we define Post-Treatment Compliers as the approximately 45% of subjects, initially assigned to either placebo or treatment condition, who answered the post-treatment survey and reported viewing the treatment videos. Table A.9 in the Appendix compares baseline measures for compliers and non-compliers at these two treatment phases. There is little evidence to suggest that the non-compliers differ significantly from the compliers. In the main text, we report the sample's intention-to-treat (ITT) results.

# Results

The experimental treatment is a short video narrative that reports the magnitude of malfeasance in a subject's municipality. Our results support the pre-registered hypothesis that corruption beliefs, measured pre-treatment (Corrupt$_{ti}$) and post-treatment (Corrupt$_{(t+1)i}$) respond to this randomly assigned information intervention.[12] Since these messages are customized to reflect actual au-

[12] We pre-registered our experimental design and analysis plan at the AEA RCT Registry [link excluded to preserve anonymity].

dit outcomes in each municipality, there is considerable variation in the quantitative details that subjects observe. We exploit this variation in order to gain some understanding of how different elements of the malfeasance message affects belief updating. While these features of the message are not randomly assigned they provide insights into how average citizens engage with malfeasance information reported in audits. We explore heterogeneous treatment effects across political partisans. And finally we present a set of results assessing the effect of the information treatment on donation and voting behavior.

## Information Treatment.

When subjects in our experiment are treated with audit agency information about local municipal government their assessment of corruption levels in their community increases quite dramatically. In our experiment, treatment, $T_i$, is randomly assigned to subjects. Treated subjects, $T_i = 1$, receive CGO information about actual corruption levels in the respondent's municipal government, i.e., $\text{Corrupt}_{ti}^{True}$. In the upper left panel of Figure 3 we observe a large significant effect for all three outcome metrics – posterior corruption beliefs for treated respondents are significantly higher, on average, than their priors while there is no significant change, on average, for participants in the placebo arm.[13] Regression coefficients for this treatment variable in Table 2 confirm this strong treatment effect. Controlling for malfeasance priors, citizens update when they have access to audit information.

**Magnitudes.** The treatment video reports the audited amount of malfeasance and also expresses these magnitudes in terms of foregone flu vaccines in order to facilitate engagement and comprehension. The bottom panel in Figure 3, presents average updating broken down by four bins of our standardized measure of malfeasance (on the x-axis). Figure 3 distinguishes between respondents living in municipalities where malfeasance increased (positive) versus those where it declined (negative). Focusing on the majority of subjects who observed an increase in reported malfeasance, we

[13] The "on average" is important here because as we see below, respondents with high versus low priors, within both these treatment arms, respond quite differently.

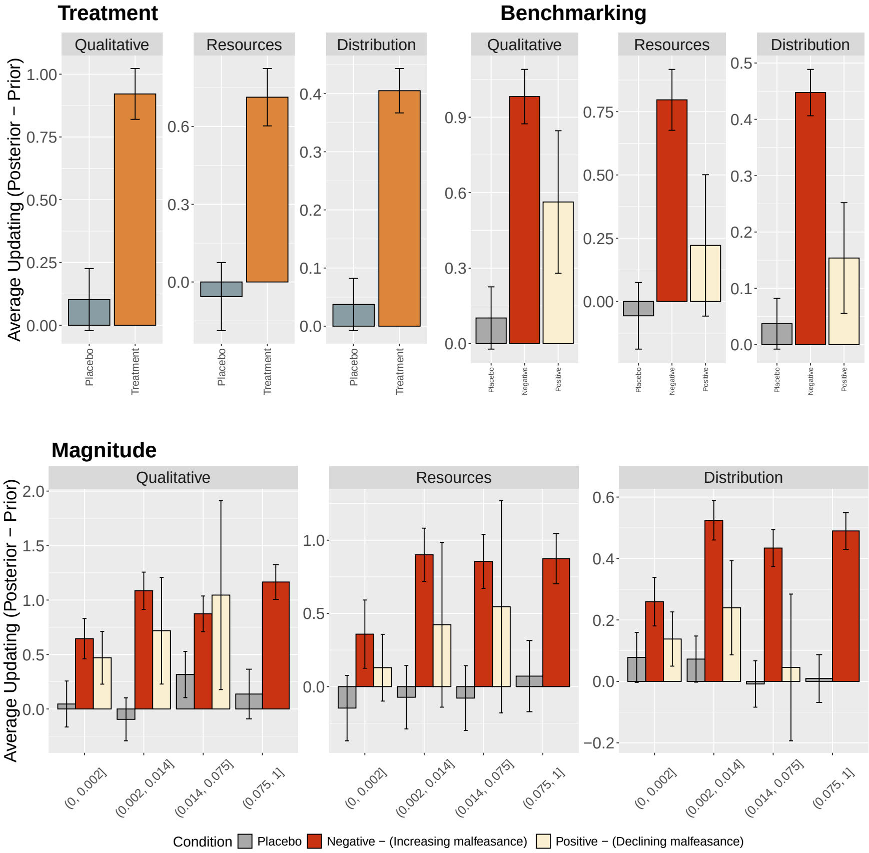
see that subjects in the lowest malfeasance quartile bin exhibited significantly lower increases in malfeasance beliefs. However, across all the other three standardized malfeasance groups, there is no statistically significant difference in average belief updating. For respondents in municipalities with improving audit reports, there is no statistically significant difference in updating across all four malfeasance magnitude bins.

For our three outcome variables, we specify a *Malfeasance Magnitude* model that includes the standardized measure of malfeasance and its interaction with the treatment variable. Table 2 reports the regression results for the full sample and also separately for municipalities reporting negative versus positive trends in their audit reports. Across all three outcome variables, we see no significant interaction between malfeasance magnitude and the treatment variable. While we have strong evidence that treated subjects believe corruption levels are higher post-intervention, there is little evidence that this belief updating is correlated with the magnitudes of the malfeasance reported in the audit information videos.

**Benchmarking.** In the second frame of the treatment video, subjects received a temporal comparison of their 2020 municipal audit results with the municipality's previous results. Of the total 116 municipalities in the study, 94 registered worse malfeasance audits in 2020 than in their previous audit; 22 had improved outcomes. We create a *Negative* subsample consisting of subjects in municipalities that experienced rising levels of malfeasance (compared to their last audit); the *Positive* subsample consists of all remaining subjects for whom malfeasance levels did not change or declined. The expectation is that beliefs about corruption should increase for treated subjects in the *Negative* subsample and would decrease for subjects in the *Positive* subsample. Beliefs are hypothesized to respond to a crude temporal comparison: whether things have gotten worse or better.

The upper right frame of Figure 3 compares subjects in municipalities with declining malfeasance audits (positive) with those in municipalities with audits that report more corruption than their previous audit (negative). On average, subjects in both conditions – those with positive and negative trends in municipal malfeasance – had higher corruption belief posteriors. Those in the positive condition, though, did exhibit significantly lower average updating of their corruption beliefs than those in the negative condition. Moreover, for the incentivized outcomes, Distribution and Resources, average updating by those in the positive audit condition was statistically no different than what was observed for the placebo subjects. Regardless of whether subjects observed positive or negative benchmarking, on average, their corruption posteriors were higher than their priors.

Figure 3: Audit Treatment Effects.



*Note:* Average belief updating (Posterior − Prior) by treatment and placebo groups, disaggregated by treatment status, benchmarking conditions, and malfeasance magnitude. The *Treatment* panel (top left) shows overall updating by group. The *Benchmarking* panel (top right) compares *Negative* and *Positive* conditions and their placebo counterparts. The *Magnitudes* panel (bottom) shows updating by treatment, split by standardized malfeasance levels.

Table 2, presents regression results for the *Negative* and *Positive* sub-samples. For all three outcome variables, the *Treat* variable has a positive coefficient for both *Negative* and *Positive* sub-

Table 2: Regression Results of Malfeasance Beliefs - All Outcomes.

| | Information | Magnitudes | Negative | Positive | Information | Magnitudes | Negative | Positive | Information | Magnitudes | Negative | Positive |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | 2.734*** | 2.717*** | 2.826*** | 2.527+ | 2.302*** | 2.274*** | 2.333*** | 2.481 | 1.122*** | 1.125*** | 1.191*** | 0.861+ |
| | (0.401) | (0.402) | (0.407) | (1.341) | (0.662) | (0.648) | (0.686) | (1.969) | (0.153) | (0.152) | (0.161) | (0.459) |
| Prior | 0.467*** | 0.464*** | 0.473*** | 0.436*** | 0.446*** | 0.444*** | 0.442*** | 0.458*** | 0.421*** | 0.415*** | 0.406*** | 0.453*** |
| | (0.020) | (0.020) | (0.022) | (0.055) | (0.018) | (0.018) | (0.019) | (0.054) | (0.024) | (0.024) | (0.024) | (0.093) |
| Treat | 0.749*** | 0.736*** | 0.773*** | 0.566** | 0.740*** | 0.739*** | 0.834*** | 0.316 | 0.347*** | 0.327*** | 0.375*** | 0.073 |
| | (0.079) | (0.091) | (0.103) | (0.203) | (0.088) | (0.094) | (0.103) | (0.200) | (0.028) | (0.035) | (0.036) | (0.081) |
| Stand - Malfeasance | | 0.490 | 0.334 | -1.877 | | 0.590 | 0.559 | 2.175 | | 0.217 | 0.220 | -9.230+ |
| | | (0.489) | (0.476) | (22.915) | | (0.432) | (0.448) | (16.192) | | (0.233) | (0.240) | (5.108) |
| Treat:Stand - Malfeasance | | 0.179 | 0.065 | 4.383 | | 0.012 | -0.259 | 11.969 | | 0.302 | 0.158 | 8.216 |
| | | (0.882) | (0.904) | (14.757) | | (0.597) | (0.518) | (16.308) | | (0.353) | (0.294) | (5.262) |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Num.Obs. | 3439 | 3439 | 2941 | 498 | 3439 | 3439 | 2941 | 498 | 3439 | 3439 | 2941 | 498 |
| R2 | 0.263 | 0.264 | 0.270 | 0.268 | 0.220 | 0.221 | 0.217 | 0.304 | 0.203 | 0.207 | 0.204 | 0.244 |
| R2 Adj. | 0.257 | 0.258 | 0.262 | 0.226 | 0.214 | 0.214 | 0.209 | 0.264 | 0.196 | 0.200 | 0.196 | 0.201 |

*Note:* Regression estimates are reported using different specifications across all three outcomes. The variable *Stand-Malfeasance* is the min-max standardized measure of the amount of malfeasance reported in the video and it ranges from 0 to 1. The *Negative* columns report the *Magnitude* specification for respondents who received negative information (increasing malfeasance), while the *Positive* columns report the specification for those who received positive information (declining malfeasance). Standard errors are clustered at the commune level. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

samples. For subjects in the *Negative* sub-sample, the *Treat* coefficient is statistically significant for all three outcome variables. In the case of the *Positive* sub-sample, the *Treat* variable is only statistically significant for the Qualitative model. Consistent with what we saw in Figure 3, treated subjects in municipalities with a deteriorating malfeasance record update more negatively than those with improving audit records. The results for the incentivized outcomes, Distribution and Resources, again consistent with Figure 3, indicate no treatment effect for subjects in the *Positive* sub-sample. Nevertheless, subjects in both the negative and positive condition have, on average, more negative posteriors than their priors.[14]

The malfeasance video treatment effect, which increases corruption belief posteriors, is significantly higher for subjects who are informed that their municipality's audit results are worse than previous audits. For subjects who are informed that audit performance has improved, we observe no significant difference from the baseline video treatment effect that increases corruption belief posteriors. As a result, all subjects, regardless of whether audits have improved or worsened, have on average higher corruption posteriors, although these are significantly higher for those in municipalities with worsening audit performance.

[14] The results for those who see the 'positive' temporal comparison should be treated cautiously. As it happens, only 22 of the municipalities had audits that improved over their previous CGO audit. This leaves us slightly less than 500 subjects in the Positive temporal category. Power calculations are provided in Appendix Section A.8.

**Information Gap.** The content of the video treatment is hypothesized to reduce gaps between ground truth and subjects' corruption beliefs $(\text{Corrupt}_{ti}^{True} - \text{Corrupt}_{ti})$. We label this gap *Corrup Diff* and generate a distinct term for both Resources and Distribution. A limitation is that pre- and post-treatment belief measures (Distribution and Resources) do not precisely reflect the malfeasance information reported in the video treatment, which consist of: 1.) the value of malfeasance expenditures and 2.) the foregone resources associated with them.[15] This variation between the signal and the updating question could attenuate our estimates of the impact of information treatments on information gaps learning (Fan et al., 2024).

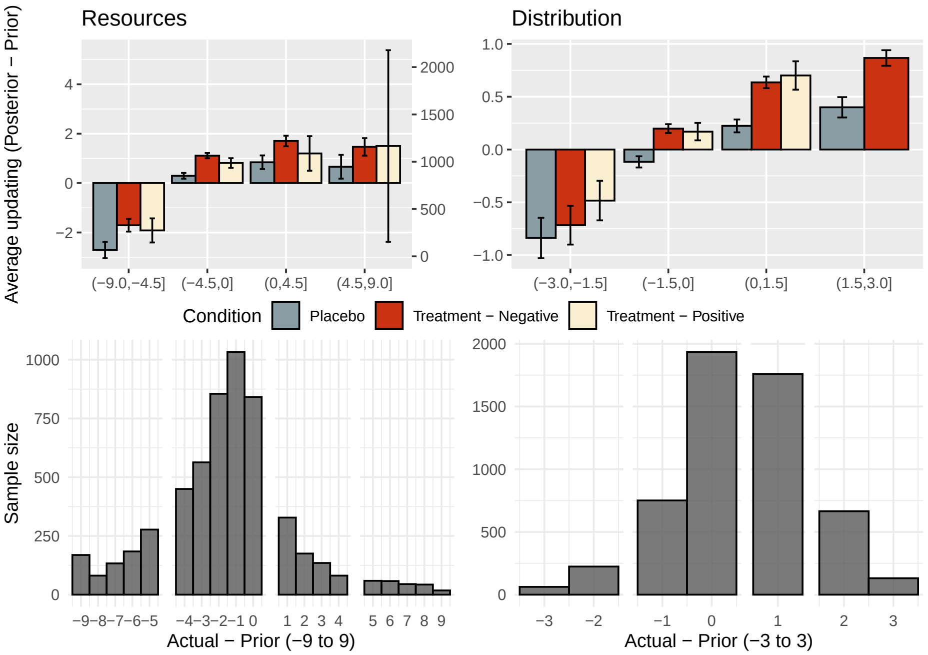
There is little evidence that belief updating reduces the subjects' information gap. The upper panel in Figure 4 presents the average updating in the y-axis ($(\text{Corrupt}_{t+1i} - \text{Corrupt}_{ti})$) for the placebo treatment arm and for those in the worsening versus improving malfeasance versions of the audit video treatments. Across all treatment arms, the gap between our corruption signal and priors in the x-axis ($(\text{Corrupt}_{ti}^{True} - \text{Corrupt}_{ti})$) is correlated with belief updating. These correlations suggest that the gap between prior beliefs and our corruption signal is not causing belief updating. One indication is the positive correlation between our variable, measuring the gap between priors and the malfeasance signal, and belief updating for subjects in the placebo treatment arm. They never receive the signal, but their updating pattern is identical to that of treated subjects who do receive the signal. The correlations we observed in Figure 4 are likely driven by regression to the mean. Those with relatively high (low) malfeasance priors register decreasing (rising) beliefs in post-treatment.

The histogram in the lower panel of Figure 4 indicates that Distribution priors are typically lower than actual corruption levels. Hence, the average for *Corrup Diff* is positive. The Resources histogram in Figure 4 suggests the opposite – they are typically higher than actual corruption levels. Hence, the average value for *Corrup Diff* is negative. On average, though, both Resource and Distribution measured malfeasance beliefs increase. As a result, for the Distribution mea-

[15] The information treatments were based on the ChatBot experiments that identified the optimal messaging strategy. These were conducted after the original pre-treatment surveys (that occurred four months before the intervention) and did not precisely match the pre-treatment belief measures.

sure, treated subjects, on average, have more accurate posterior beliefs – from an average 0.36 discrepancy pre-treatment to essentially zero in post-treatment. On the other hand, results for the Resources outcome variable suggests that subjects, on average, have less accurate beliefs about corruption in post-treatment – their accuracy goes from -1.61 to -2.22. Given that these two metrics have differently signed net belief accuracies in pre-treatment, our information gap model of updating predicts that, on average, Resource and Distribution beliefs would move in the opposite direction. In fact, they move in the same direction, and as a result, the Resource posteriors are less accurate than priors. Subjects, on average, respond to treatment (regardless of quantitative content) by worsening their assessment of corruption levels. Subjects do not appear to be using the quantitative content of the message to reduce their information gap.

Figure 4: Belief Updating by Gap Between Actual Malfeasance and Pre-Treatment Beliefs.



*Note:* The upper left panel plots average belief updating (Posterior − Prior) on the y-axis against the prior–truth gap (Corrupt$ti^{True}$ – Corrupt$ti$) on the x-axis. For the *Resources* outcome, this gap ranges from $-9$ to $+9$; for *Distribution*, from $-3$ to $+3$. Negative values indicate overestimation of corruption; positive values indicate underestimation. Treatment respondents are split into *Negative* (told corruption increased) and *Positive* (told it decreased) conditions. The lower panels display the distribution of the prior–truth gap for each outcome: bottom left for *Resources*, bottom right for *Distribution*. Related regression estimates appear in Table A.10 in the Appendix.

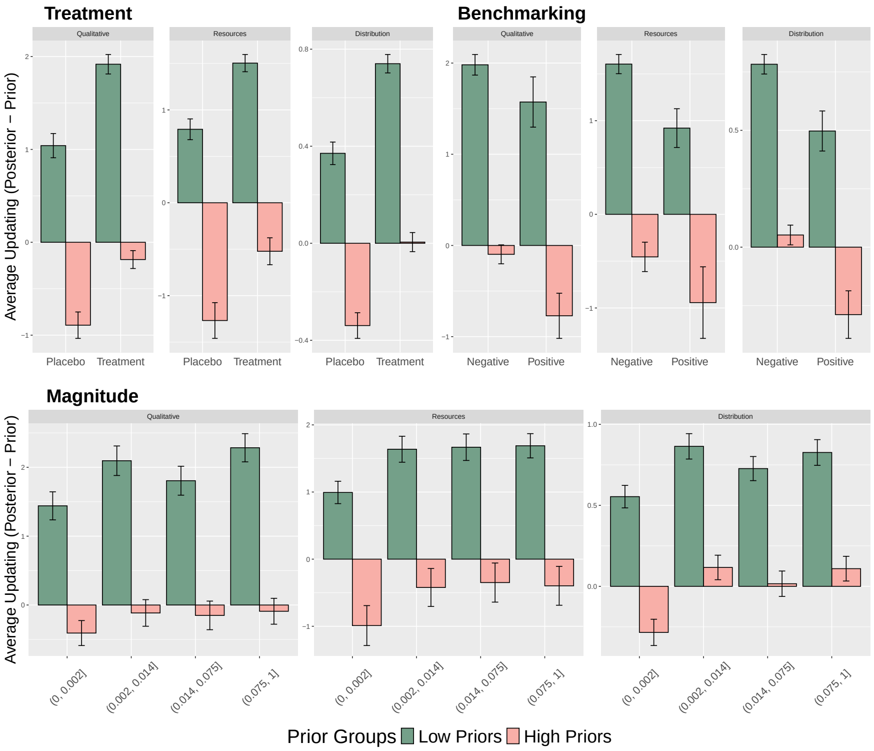
Regression estimates in Table A.10 in the Appendix confirm this conclusion. For the Distribution outcome, the estimated coefficient for the interaction term, *Treat x Corrup Diff*, is positive and precisely estimated. This could be evidence of updating that narrows an information gap. But the results for the Resources outcome suggest regression to the mean. In the Resources model, all subjects update conditional on the Corrup Diff term (including placebo subjects who receive no information signal). Note that the estimated coefficient for the interaction term, *Treat x Corrup Diff* in the Resource model, is essentially zero. The difference between actual corruption levels and subjects' priors is clearly correlated with belief updating in the Distribution model – but the overall evidence does not suggest updating aimed at reducing an information gap.

As our results suggest, and as many others have pointed out (Gill & Walker, 2005; Haaland et al., 2023; Roth & Wohlfart, 2020), it is extremely challenging to design information campaigns that reduce the difference between some ground truth and beliefs. It is also very difficult to measure this gap between beliefs and ground truth. A concern is whether we adopted the right measurement strategy for determining whether posterior beliefs are better aligned with ground truth after our treatment intervention. In recognition of these challenges, we adopted multiple metrics of this gap between ground truth and beliefs; we incentivized the elicitation of these beliefs; and we implemented a within subject pre- and post-design. If posterior beliefs were better aligned with ground truth then we should have detected evidence of this with our design. Contradictory trends in our multiple belief measures and regression to the mean for subjects in both treatment and placebo lend little support for the notion that information reduces the gap between ground truth and beliefs.

**High/Low Malfeasance Priors.** For each of the three malfeasance outcome variables, we grouped subjects into those scoring lower versus higher than the median value for their municipal sample (a low/high score indicates subjects believe malfeasance is low/high). The upper left *Treatment* frame of Figure 5 indicates that the direction of updating is conditional on whether subjects had high or low priors. Respondents in the placebo condition receive no new information about malfeasance in their municipality. Yet, for all three outcome variables, placebo subjects with high priors exhibit significant declining beliefs regarding malfeasance, and those with low priors exhibit the opposite pattern. In the absence of the treatment video, the updating is roughly symmetric for those with high and low priors.

The pattern is quite different for subjects receiving the malfeasance audit treatment. There is a clear pattern of asymmetric belief updating. First, across all three outcome variables we observe much higher levels of belief updating for those with low priors than those with high priors. Also note in the upper left frame that belief updating by subjects in treatment and control differ considerably: Treated subjects with low priors update more negatively compared to the low prior placebo subjects and conversely the treated subjects with high priors update less positively than high prior subjects in placebo.

Figure 5: Audit Treatment Updating Effects by Low versus High Priors.



*Note:* This figure shows average belief updating, measured as the difference between posteriors and priors (Posterior − Prior), conditional on treatment status, benchmarking, and standardized malfeasance levels. Within each group, respondents are further split by prior beliefs: *Low* (below-median priors within their municipality) and *High* (above-median). The top left panel (Treatment) presents updating for treated and placebo groups, disaggregated by prior level. The top right panel (Benchmarking) shows treated respondents only, split by benchmark type (*Negative* or *Positive*) and prior level. The bottom panel (Magnitudes) displays average updating for treated respondents by standardized malfeasance levels and prior group.

We observe a similar pattern in the upper right *Benchmarking* panel of Figure 5 that compares belief updating for subjects receiving positive versus negative benchmarked audit results. Updating is broken down by low/high priors. Subjects with low priors regarding malfeasance, regardless of whether they received information about improving or declining levels of malfeasance in their municipal government, registered significant worsening beliefs regarding malfeasance. On the other hand, subjects with high malfeasance priors update at a much lower rate than subjects with low priors. And while clearly there is asymmetry, there is a benchmarking effect: Subjects who received positive benchmarked audits update less negatively if they had low priors and more positively if they had high priors.

The bottom *Magnitude* frame of Figure 5 presents average belief updating for those with low and high priors organized by the standardized measure of malfeasance reported for their municipality. Again, we observe similar asymmetries between those with high and low priors. Those with low priors, regardless of their malfeasance magnitude grouping, have significant increasing average beliefs about corruption levels. Subjects who registered high corruption priors respond in a much more muted fashion to the malfeasance video treatment.

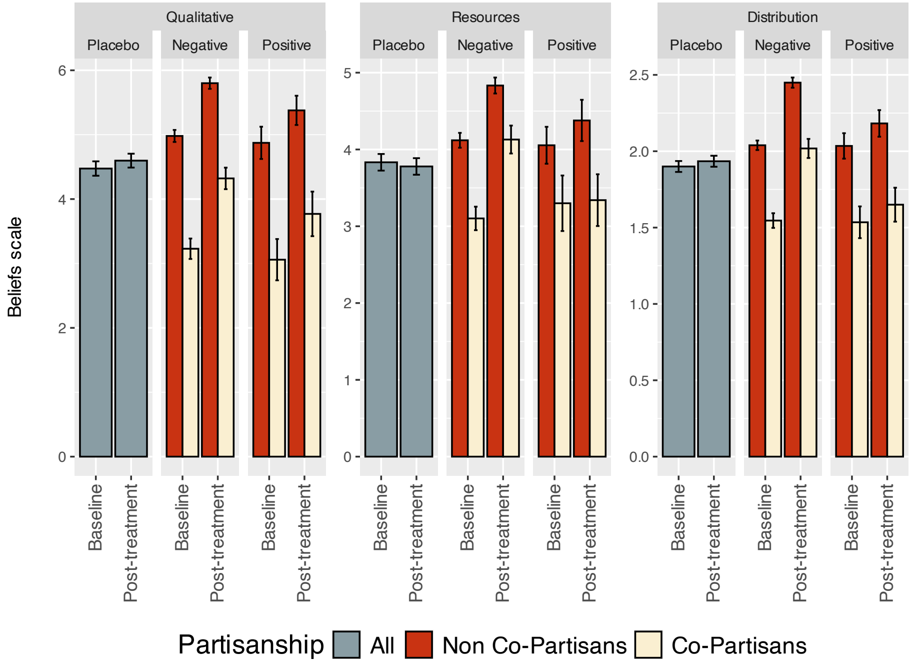
The asymmetric belief updating observed in Figure 5 suggests that our estimated treatment effects vary by whether subjects had high versus low priors. Table A.11 in the Appendix presents the low/high regression results for four model specifications: *Treatment*, *Benchmarking*, *Magnitudes* and *Information Gap*. In each case, we report results for the total sample, those with low priors and respondents with high priors. For all outcome variables, the coefficient on *Treat* is very similar for the Pooled, High Prior, and Low Prior samples. And the null effects we reported for the *Magnitudes* and *Information Gap* models are confirmed when we segment the sample into subjects with high and low priors. The exception is the *Benchmarking* model in Table A.11: For the full sample, informing subjects that their municipal audit was worse then previous audits significantly increased their corruption beliefs. The breakouts in Table A.11 indicate that this benchmarking effect is primarily driven by those with low priors. Subjects who, pre-treatment, believe corruption levels are high are not affected by learning that their municipality's audit has worsened.

## Partisanship

Our experimental results are consistent with the notion of parallel persuasion: There are co-partisan differences in corruption priors, but co-partisans and non co-partisans respond similarly to quantitative information. Figure 6 illustrates comparisons of mean corruption beliefs for co-partisans (yellow bars) and non co-partisans (red bars); pre-treatment and post-treatment. Respondents are separated into those who received a negative versus positive corruption audit report. Focusing simply on non-co-partisan subjects (the red bars) and co-partisan subjects (the yellow bars), we observe that non co-partisans consistently have higher beliefs about levels of municipal government corruption than is the case for partisans of the local government. This holds for all three pre-treatment and post-treatment outcome variables.

Figure 6 also allows for a comparison of the audit video treatment effects for partisans and non co-partisans. The pattern of treatment effects is similar for partisans and non-co-partisans across all three outcome variables. For the Qualitative outcome variable, average updating responds similarly for both partisan groups, which holds for both positive and negative versions of the information treatment. Both partisan groups respond similarly to the negative audit information treatment for the Distribution and Resources outcomes. Regarding the positive information treatment, partisans and non co-partisans register non-significant treatment effects.

Figure 6: Corruption Priors and Posteriors Co-Partisans and Non-Co-partisans.



*Note:* This figure shows prior and posterior beliefs by treatment and partisanship. The sample is split into "Co-Partisans" (those intending to vote for the incumbent) and "Non Co-Partisans" (those not supporting the incumbent). Treatment groups include Placebo, Negative (increasing malfeasance), and Positive (declining malfeasance).

In Table 3, we confirm this notion of parallel persuasion by re-estimating the regression models in Table 2 separately for the co-partisans and non co-partisans. The *Treatment* effect for all three outcome variables is large, significant, and similar for both co-partisans and non co-partisans. Concerning the *Benchmarking* model, the net effect of the negative treatment is similar for co-partisans and non co-partisans across all three outcome variables. Similarly, there are minimal partisan differences in the *Magnitudes* model: treatment effect across the three outcome measures are similar for both partisan groups. Co-partisans have significantly more positive priors regarding levels of malfeasance in their municipal government. But partisan preferences have a relatively limited impact on how subjects update their malfeasance beliefs.

Table 3: Regression Results of Malfeasance Beliefs - Co-Partisan and Non Co-partisans

| | Qualitative | | Resources | | Distribution | |
|---|---|---|---|---|---|---|
| | Co-partisan | Non-copartisan | Co-partisan | Non-copartisan | Co-partisan | Non-copartisan |
| **Panel A: Information** | | | | | | |
| Intercept | 2.061+ | 3.144*** | 2.877 | 2.125*** | 1.058** | 1.166*** |
| | (1.115) | (0.393) | (1.885) | (0.533) | (0.401) | (0.153) |
| Prior | 0.424*** | 0.415*** | 0.414*** | 0.435*** | 0.382*** | 0.381*** |
| | (0.037) | (0.023) | (0.038) | (0.019) | (0.050) | (0.026) |
| Treat | 0.742*** | 0.733*** | 0.788*** | 0.723*** | 0.303*** | 0.357*** |
| | (0.152) | (0.083) | (0.135) | (0.107) | (0.056) | (0.034) |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| Num.Obs. | 929 | 2510 | 929 | 2510 | 929 | 2510 |
| R2 | 0.195 | 0.233 | 0.219 | 0.215 | 0.158 | 0.193 |
| R2 Adj. | 0.172 | 0.224 | 0.197 | 0.206 | 0.134 | 0.185 |
| **Panel B: Benchmarking** | | | | | | |
| Intercept | 1.483 | 3.121*** | 2.471 | 2.236*** | 0.926* | 1.223*** |
| | (1.151) | (0.417) | (1.864) | (0.517) | (0.388) | (0.178) |
| Prior | 0.424*** | 0.416*** | 0.414*** | 0.435*** | 0.377*** | 0.382*** |
| | (0.036) | (0.022) | (0.038) | (0.020) | (0.048) | (0.028) |
| Treat | 0.908 | 0.362** | 0.471* | 0.213 | 0.143 | 0.080 |
| | (0.498) | (0.139) | (0.186) | (0.251) | (0.100) | (0.101) |
| Negative | 0.710* | 0.036 | 0.494* | -0.106 | 0.169 | -0.055 |
| | (0.292) | (0.234) | (0.207) | (0.246) | (0.106) | (0.113) |
| Treat:Negative | -0.217 | 0.430* | 0.371 | 0.590* | 0.189 | 0.320** |
| | (0.528) | (0.170) | (0.247) | (0.275) | (0.120) | (0.107) |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| Num.Obs. | 929 | 2510 | 929 | 2510 | 929 | 2510 |
| R2 | 0.204 | 0.236 | 0.231 | 0.217 | 0.175 | 0.201 |
| R2 Adj. | 0.179 | 0.227 | 0.207 | 0.208 | 0.149 | 0.192 |
| **Panel C: Magnitudes** | | | | | | |
| Intercept | 1.932+ | 3.142*** | 2.651 | 2.126*** | 1.011** | 1.172*** |
| | (1.084) | (0.395) | (1.789) | (0.527) | (0.363) | (0.155) |
| Prior | 0.417*** | 0.415*** | 0.412*** | 0.434*** | 0.368*** | 0.378*** |
| | (0.038) | (0.022) | (0.038) | (0.019) | (0.049) | (0.027) |
| Treat | 0.783*** | 0.713*** | 0.884*** | 0.698*** | 0.314*** | 0.330*** |
| | (0.180) | (0.092) | (0.160) | (0.115) | (0.068) | (0.041) |
| Stand - Malfeasance | 1.844 | 0.144 | 2.536 | 0.121 | 0.819+ | 0.081 |
| | (1.991) | (0.394) | (1.903) | (0.294) | (0.486) | (0.274) |
| Treat:Stand - Malfeasance | -0.907 | 0.303 | -1.846 | 0.375 | -0.277 | 0.394 |
| | (2.135) | (0.673) | (2.450) | (0.465) | (0.664) | (0.320) |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| Num.Obs. | 929 | 2510 | 929 | 2510 | 929 | 2510 |
| R2 | 0.199 | 0.233 | 0.225 | 0.215 | 0.166 | 0.197 |
| R2 Adj. | 0.174 | 0.224 | 0.200 | 0.206 | 0.140 | 0.188 |
| **Panel D: Information Gap** | | | | | | |
| Intercept | | | 2.713 | 2.015*** | 0.766* | 1.071*** |
| | | | (1.846) | (0.530) | (0.351) | (0.162) |
| Prior | | | 0.507*** | 0.476*** | 0.468*** | 0.428*** |
| | | | (0.041) | (0.027) | (0.064) | (0.035) |
| Treat | | | 0.739*** | 0.741*** | 0.332*** | 0.333*** |
| | | | (0.146) | (0.120) | (0.076) | (0.034) |
| Corrup diff | | | 0.122* | 0.034 | 0.129* | -0.006 |
| | | | (0.049) | (0.033) | (0.053) | (0.035) |
| Treat:Corrup diff | | | -0.038 | 0.012 | -0.045 | 0.085** |
| | | | (0.053) | (0.033) | (0.058) | (0.026) |
| Covariates | | | Yes | Yes | Yes | Yes |
| Num.Obs. | | | 929 | 2510 | 929 | 2510 |
| R2 | | | 0.226 | 0.216 | 0.167 | 0.199 |
| R2 Adj. | | | 0.202 | 0.207 | 0.141 | 0.190 |

*Note:* This table replicates the models from Tables 2 and A.10 in the Appendix, disaggregated by partisanship (”Co-partisans” vs. ”Non-co-partisans”). Co-partisans are respondents who reported an intention to vote for the incumbent in the upcoming municipal election, while non-co-partisans include those undecided or intending to vote against the incumbent. Voting preferences were measured in the pre-treatment, post-treatment, and follow-up surveys. In the Information Gap model, *Corrup Diff*, refers to the difference between respondents' priors and the true level of malfeasance: $(\text{Corrupt}_{ti}^{True} - \text{Corrupt}_{ti})$ Standard errors are clustered at the commune level. ${}^{*}p < .05$, ${}^{**}p < .01$, ${}^{***}p < .001$.

## Spurious or Genuine Updating.

Belief updating was measured one-week and then one-month after the information treatment, providing strong evidence that the information treatment effect persists.[16] Figure A.5 in the Appendix summarizes average updating, for all three outcome variables, one-week and one-month follow-up. We present persistence ratios by dividing one-month estimates by one-week estimates, For overall treatment effects, the estimated persistent ratios range between 0.4 and 0.71. Treatment effects persist, particularly for respondents who received a negative benchmark information treatment, yielding ratios between 0.5 and 0.73. However, these ratios considerably decline for the positive benchmark treatment, ranging from 0 to 0.35.[17]

## Behavior

An important question, for which we can only provide speculative evidence here, is whether the updated beliefs we observe affect political behaviors. We evaluate the correlation between belief updating and two behavioral measures, both measured pre- and post-treatment: a costly donation to municipal public goods and vote intention for the incumbent. We make no causal claims here because the belief updating is not randomly assigned.

**Donating.** In both pre-and post-treatment, we give subjects the opportunity to make a 500 CLP (around 15% of their subject payments) contribution to their local municipality.[18] We treat this as an indicator of our subjects' willingness to contribute to the cost of providing local public goods (Beekman et al., 2014; Bonica, 2018). For each of our three outcome variables, we distinguish between those subjects that updated favorably regarding malfeasance versus those updating unfavorably.

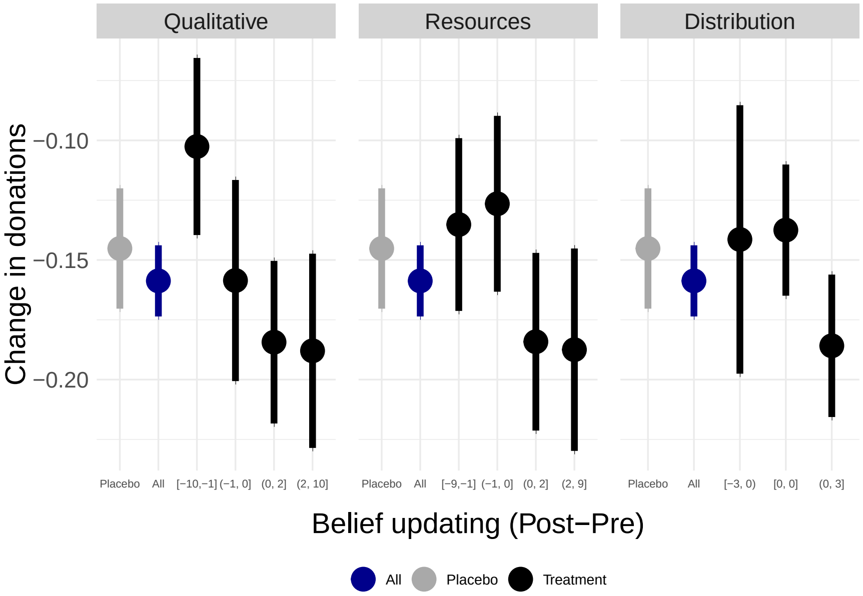
[16] Recent efforts to measure the persistence of information treatments implement additional post-treatment measurements that typically range between four and eight weeks (Cavallo et al., 2017; Haaland & Roth, 2021; Kuziemko et al., 2015).

[17] We even find that average updating estimates switched from positive to negative updating for the Resources outcome, when the treatment is positive.

[18] 500 CLP is around 50 cents.

Figure 7 presents the correlations between belief updating and change in donation behavior. The x-axis organizes changes in belief updating (posterior - prior) into quartiles for those in the audit information treatment arm. While the y-axis captures change in donation behavior (post-treatment minus pre-treatment). On average, all subjects in both treatment and placebo conditions donated less post-treatment. In the case of the Qualitative outcome measure, there is evidence that those who updated most positively regarding municipal government malfeasance had the smallest reduction in donations post-treatment. There is no evidence of a correlation in the case of the other two incentivized outcome measures. On balance, there is little evidence in Figure 7 to suggest that the belief updating we measure is correlated with a costly donation to municipal services.

Figure 7: Change in Donations - Belief Updating.

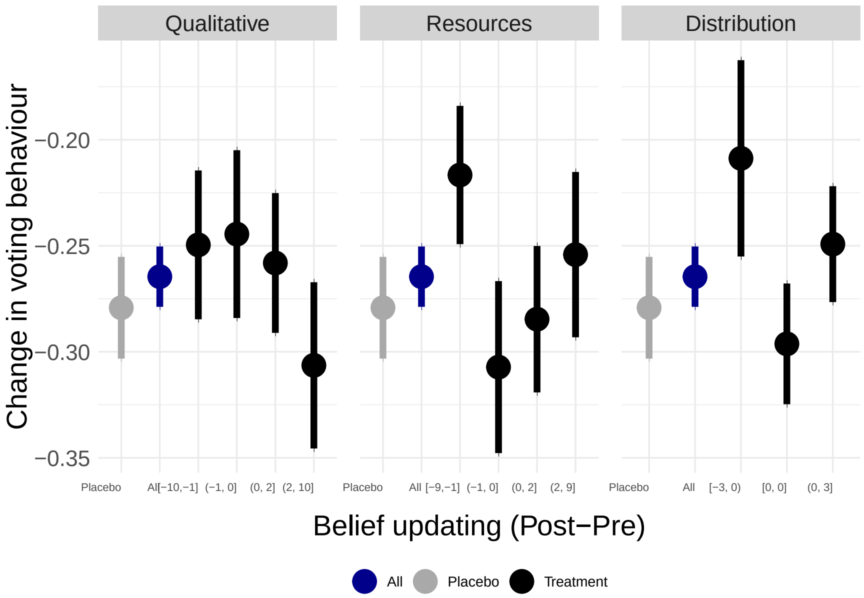


*Note:* This figure shows the average change in respondents' decisions to donate to their municipality from baseline to post-treatment (y-axis) with 95% confidence intervals. Changes are conditional on belief updating across all three outcome measures (x-axis), divided into quartiles and with the full sample average in blue. Donations were coded as 1 (donated) or 0 (did not). Positive changes indicate new donations post-treatment; negative changes indicate stopping donations.

**Voting.** Both pre- and post-treatment we asked subjects to indicate whether they would vote for the incumbent mayor. The y-axis of Figure 8 represents the change in the percent of subjects voting for the incumbent (post-treatment minus pre-treatment). The x-axis is equivalent to the quartiles and terciles presented in Figure 7. On average, considering all subjects, the percentage of expressing an incumbent vote preferences drops by about 27% and we see a similar average decline for subjects in the placebo arm. The expectation is that these declines in average incumbent vote intention would become larger as corruption belief updating (post-treatment minus pre-treatment) increased. There is very little evidence that this is the case. In the case of the incentivized belief measures, for those who updated most positively about corruption levels, we do see declines in incumbent support that are significantly lower than those in the placebo treatment arm. But otherwise the changes in voting behavior are similar across the belief updating categories.

The malfeasance belief updating we observe post-treatment has very little behavioral consequences. We implemented two very distinct behavioral outcome measures: costly contributions to a municipal public good; and vote intention for the incumbent mayor. Favorable updating of malfeasance beliefs has no significant effect on either donations behavior nor on incumbent vote intention.

Figure 8: Voting Intention Conditional on Belief Updating.



*Note:* This figure shows the average change in respondents' intention to vote for the incumbent mayor from baseline to post-treatment (y-axis) with 95% confidence intervals. Estimates are conditional on belief updating across three outcome measures (x-axis), disaggregated into quartiles; the full-sample average is shown in *blue*. Responses were coded 1 for intending to vote for the incumbent, 0 otherwise. A positive change $(1 - 0)$ indicates a shift from not supporting to supporting the incumbent, and a negative change $(0 - 1)$ indicates the reverse.

# Discussion and Conclusions

Our study examines how individuals' beliefs about corruption respond to quantitative information regarding actual governmental malfeasance in their community. We implement a field experiment in Chile that incorporates the 116 CGO municipal audit results from 2020 into video information treatments that are randomly assigned to 5,528 online subjects. Based on extensive piloting, the information treatment presented the audited total costing of the malfeasance; how this compared to previous audits; and foregone public expenditures resulting from the malfeasance. Beliefs about malfeasance are measured with three different variables: a qualitative variable (non-incentivized) and two incentivized quantitative variables.

Subjects responded to corruption information. Receiving a video message summarizing the CGO's audit of irregular activities in their municipality causes Chileans to update their beliefs about corruption in their municipality. This estimated treatment effect is very robust. For all three outcome variables the treatment effect is large and precisely estimated. This is strong evidence that beliefs about malfeasance levels rise when average citizens are informed about corruption in their municipal government. This effect was more pronounced among those with initially low assessments of government malfeasance.

The quantitative content of the corruption video treatments varied by municipality. The video provided two pieces of information regarding malfeasance magnitude: the total value of audited malfeasance and this amount expressed in terms of foregone municipal services. This content was not randomly assigned. We find little evidence that subjects' belief updating correlates with the magnitudes of municipal government malfeasance that is reported in the audit reports. Respondents who were informed that their municipality's audit was worse than previous audits exhibited a significantly larger increase in their corruption beliefs. Overall, subjects in municipalities with both improving and deteriorating audit performance had, on average, more negative posteriors than their priors.

We conjectured that treated subjects' posterior beliefs would better align with ground truth than was the case for their prior beliefs. The evidence for our two quantitative outcome variables is contradictory. Subjects' priors under-estimated corruption levels measured by the Distribution metric and over-estimated levels in the case of the Resources metric. For both metrics we observe higher average posterior levels of malfeasance beliefs. As a result we observe a narrowing of the gap between priors and ground truth for the Distribution metric but a widening gap for the Resources metric. Regardless of the gap between their priors and ground truth, most treated subjects have higher corruption posteriors. Subjects are not using the information treatment to calibrate their posteriors so they better align with ground truth.

Partisanship does not significantly undermine the causal effect of malfeasance information on belief updating. Co-partisans of the incumbent mayor have less negative beliefs than non co-partisans about levels of malfeasance. But both segments of the sampled population respond similarly to being treated by a malfeasance information video.[19]

Finally, we do not find evidence that corruption belief updating affects political behavior. There is no evidence that a subjects' willingness to donate to a municipal public good in the post-treatment survey (controlling for their pre-treatment donation decision) is correlated with their observed belief updating. Similarly, post-treatment intention to vote for the incumbent is uncorrelated with updating of corruption beliefs.

We conclude that malfeasance information messages that reach the public cause belief updating; updating tends to raise the average level of posterior beliefs about municipal government malfeasance. Subjects received corruption messaging customized to their particular municipality. These customized messages provided details regarding magnitudes and relative performance compared to previous audits but they did not necessarily result in significantly more ”informed” corruption beliefs.

**Implications** Treated participants in our experiment came away with the broad impression that their local government engaged in wrongdoing, but fewer appeared to internalize the specifics of

[19] For two of the outcome variables, subjects received financial incentivizes for correctly estimating malfeasance levels. As others have pointed out this may have played a role in moderating partisan bias in judgments (Rathje et al., 2023).

how much, compared to what, and compared to when. This illustrates a broader point: citizens often struggle with complex numerical data and nuanced policy metrics. Cognitive constraints and limited attention lead people to rely on heuristics or simplified interpretations when confronted with detailed statistics.[20] A one-time exposure, even when packaged in an accessible format, may be insufficient to overcome these limitations. It might be unrealistic, as our evidence intimates, to expect a short informational treatment to significantly improve public understanding of quantitative audit findings.

Addressing this challenge will likely require innovation in both content and delivery of anti-corruption information. One direction is to simplify and contextualize key facts even further – for instance, using vivid visual aids, analogies, or personalized impact statements to drive points home. Another approach is repetition and reinforcement: citizens may need to be exposed to important information multiple times or through multiple channels before it truly sinks in. Encouraging public deliberation or discussion around the information could also deepen engagement, as talking through the implications may prompt individuals to process details more carefully than passive listening would. Recent research has warned that poorly executed transparency efforts can falter or even produce adverse effects (Cheeseman & Peiffer, 2020; Chong et al., 2015). Our study adds to this caution, highlighting that well-intentioned information campaigns must grapple with the reality of superficial processing.

Our results offer one potential explanation for the mixed record of information-based accountability interventions observed in prior research. Field experiments that have provided voters with audit findings or performance scorecards often report modest or null effects on electoral outcomes (Dunning, Grossman, Humphreys, Hyde, et al., 2019). If citizens only shallowly process the information – updating their general beliefs about "corruption in government" but not differentiating by degree or assigning responsibility carefully – then we should not be surprised that their voting behavior does not dramatically change.

Partisanship is often singled out as undermining the public's acceptance of negative informa-

[20] In Section A.5, we explore heterogeneity using Bayesian Additive Regression Trees to identify whether respondents with higher levels of education respond differently to subjects with low levels of education.

tion. We observe though, that factual evidence of malfeasance increased corruption beliefs for both allies and opponents of the local government. This result is encouraging: it implies that credible, concrete information can cut through partisan filters, at least in terms of belief updating. The design of our study – including non-partisan framing of the audit and financial incentives for accurate answers – may have helped mitigate motivated reasoning. Our contribution here is to show that even though partisanship shapes baseline perceptions of corruption, it does not prevent citizens from updating their beliefs when presented with hard evidence. This nuance adds a hopeful note to the otherwise cautionary tale of limited information processing.

Our study suggests that the accountability gains from transparency are limited unless citizens can be helped to engage with details. Simply put, making government data public is not the same as making it meaningful to the public. Strengthening democratic accountability requires closing this gap between information availability and citizens' understanding.

# References

Adida, C., Gottlieb, J., Kramon, E., & McClendon, G. (2020). When does information influence voters? the joint importance of salience and coordination. *Comparative Political Studies*, *53*(6), 851–891.

Anderson, K., Zamarro, G., Steele, J., & Miller, T. (2021). Comparing performance of methods to deal with differential attrition in randomized experimental evaluations. *Evaluation Review*, *45*(1-2), 70–104.

Anduiza, E., Gallego, A., & Muñoz, J. (2013). Turning a blind eye: Experimental evidence of partisan bias in attitudes toward corruption. *Comparative Political Studies*, *46*(12), 1664–1692.

Arel-Bundock, V., Blais, A., & Dassonneville, R. (2021). Do voters benchmark economic performance? *British Journal of Political Science*, *51*(1), 437–449. https://doi.org/10.1017/S0007123418000236

Arias, E., Balán, P., Larreguy, H., Marshall, J., & Querubín, P. (2019). Information provision, voter coordination, and electoral accountability: Evidence from mexican social networks. *American Political Science Review*, *113*(2), 475–498. https://doi.org/10.1017/S0003055419000091

Arias, E., Larreguy, H., Marshall, J., & Querubín, P. (2022a). Priors Rule: When Do Malfeasance Revelations Help Or Hurt Incumbent Parties? *Journal of the European Economic Association*, *20*(4), 1433–1477.

Arias, E., Larreguy, H., Marshall, J., & Querubín, P. (2022b). Priors rule: When do malfeasance revelations help or hurt incumbent parties? *Journal of the European Economic Association*, *20*(4), 1433–1477. https://doi.org/10.1093/jeea/jvac015

Arias, E., Larreguy, H. A., Marshall, J., & Querubín, P. (2025). Does the content and mode of delivery of information matter for electoral accountability? evidence from a field experiment in mexico. *Latin American Economic Review*, *34*, 1–41. https://doi.org/10.60758/laer.v34i.335

Avenburg, A. (2019). Public costs versus private gain: Assessing the effect of different types of information about corruption incidents on electoral accountability. *Journal of Politics in Latin America*, *11*(1), 71–108. https://doi.org/10.1177/1866802X19840457

Aytaç, S. E. (2018). Relative economic performance and the incumbent vote: A reference point theory. *The Journal of Politics*, *80*(1), 16–29. https://doi.org/10.1086/693908

Bandiera, O., Prat, A., & Valletti, T. (2009). Active and passive waste in government spending: Evidence from a policy experiment. *American Economic Review*, *99*(4), 1278–1308. https://doi.org/10.1257/aer.99.4.1278

Banerjee, A., Green, D. P., McManus, J., & Pande, R. (2014). Are poor voters indifferent to whether elected leaders are criminal or corrupt? a vignette experiment in rural india. *Political Communication*, *31*(3), 391–407. https://doi.org/10.1080/10584609.2014.914615

Barro, R. (1973). The control of politicians: An economic model. *Public Choice*, *14*(1), 19–42. https://doi.org/10.1007/BF01718440

Bauhr, M., & Charron, N. (2018). Insider or outsider? grand corruption and electoral accountability. *Comparative Political Studies*, *51*(4), 415–446. https://doi.org/10.1177/0010414017710258

Becher, M., Brouard, S., & Stegmueller, D. (2023). Endogenous benchmarking and government accountability: Experimental evidence from the covid-19 pandemic. *British Journal of Political Science*, 1–18. https://doi.org/10.1017/S0007123423000170

Beekman, G., Bulte, E., & Nillesen, E. (2014). Corruption, investments and contributions to public goods: Experimental evidence from rural liberia. *Journal of Public Economics*, *115*, 37–47.

Berliner, D., & Wehner, J. (2022). Audits for accountability: Evidence from municipal by-elections in south africa. *The Journal of Politics*, *84*(3), 1581–1594. https://doi.org/10.1086/716951

Bhandari, A., Larreguy, H., & Marshall, J. (2019). *Able and mostly willing: An empirical anatomy of information's effect on voter-driven accountability in 7yhb77uhn mvk,87-* [Working paper].

Blinder, S., & Schaffner, B. F. (2019). Going with the flows: Information that changes americans' immigration preferences. *International Journal of Public Opinion Research*, *32*(1), 153–164. https://doi.org/10.1093/ijpor/edz007

Boas, T. C., Hidalgo, F. D., & Melo, M. A. (2019). Norms versus action: Why voters fail to sanction malfeasance in brazil. *American Journal of Political Science*, *63*(2), 385–400. https://doi.org/10.1111/ajps.12413

Bobonis, G. J., Cámara Fuertes, L. R., & Schwabe, R. (2016). Monitoring corruptible politicians. *American Economic Review*, *106*(8), 2371–2405. https://doi.org/10.1257/aer.20130874

Bonica, A. (2018). Are donation-based measures of ideology valid predictors of individual-level policy preferences? *The Journal of Politics*, *81*. https://doi.org/10.1086/700722

Botero, S., Cornejo, R. C., Gamboa, L., Pavao, N., & Nickerson, D. W. (2015). Says who? an experiment on allegations of corruption and credibility of sources. *Political Research Quarterly*, *68*(3), 493–504. https://doi.org/10.1177/1065912915591607

Breitenstein, S. (2019). Choosing the crook: A conjoint experiment on voting for corrupt politicians. *Research & Politics*, *6*(1), 2053168019832230. https://doi.org/10.1177/2053168019832230

Brollo, F., Nannicini, T., Perotti, R., & Tabellini, G. (2013). The political resource curse. *American Economic Review*, *103*(5), 1759–96. https://doi.org/10.1257/aer.103.5.1759

Buckley, N., Reuter, O. J., Rochlitz, M., & Aisin, A. (2022). Staying out of trouble: Criminal cases against russian mayors. *Comparative Political Studies*, *55*(9), 1539–1568. https://doi.org/10.1177/00104140211047399

Buntaine, M. T., Jablonski, R., Nielson, D. L., & Pickering, P. M. (2018). Sms texts on corruption help ugandan voters hold elected councillors accountable at the polls. *Proceedings of the National Academy of Sciences*, *115*(26), 6668–6673. https://doi.org/10.1073/pnas.1722306115

Campos-Vazquez, R. M., & Mejia, L. A. (2016). Does corruption affect cooperation? a laboratory experiment. *Latin American Economic Review*, *25*(1), 5. https://doi.org/10.1007/s40503-016-0035-0

Carey, J. M., Chun, E., Cook, A., Fogarty, B. J., Jacoby, L., Nyhan, B., Reifler, J., & Sweeney, L. (2025). The narrow reach of targeted corrections: No impact on broader beliefs about election integrity. *Political Behavior*, *47*(2), 737–750. https://doi.org/10.1007/s11109-024-09968-0

Cavallo, A., Cruces, G., & Perez-Truglia, R. (2017). Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics*, *9*(3), 1–35. https://doi.org/10.1257/mac.20150147

Charbonneau, É., & Van Ryzin, G. G. (2015). Benchmarks and citizen judgments of local government performance: Findings from a survey experiment. *Public Management Review*, *17*(2), 288–304. https://doi.org/10.1080/14719037.2013.798027

Cheeseman, N., & Peiffer, C. (2020). *The unintended consequences of anti-corruption messaging in nigeria: Why pessimists are always disappointed* (tech. rep.). Anti-Corruption Evidence SOAS Consortium.

Cheeseman, N., & Peiffer, C. (2021). The curse of good intentions: Why anticorruption messaging can encourage bribery. *American Political Science Review*, 1–15. https://doi.org/10.1017/S0003055421001398

Chesseman, N., & Peiffer, C. (2024). Anti- corruption awareness raising campaigns: Why do they fail, and how can "backfire" effects be avoided? In *Handbook of anti-corruption research and practice*. Routledge.

Chong, A., De La O, A. L., Karlan, D., & Wantchekon, L. (2015). Does corruption information inspire the fight or quash the hope? a field experiment in mexico on voter turnout, choice, and party identification. *The Journal of Politics*, *77*(1), 55–71.

Coibion, O., Georgarakos, D., Gorodnichenko, Y., & van Rooij, M. (2023). How does consumption respond to news about inflation? field evidence from a randomized control trial. *American Economic Journal: Macroeconomics*, *15*(3), 109–52. https://doi.org/10.1257/mac.20200445

Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, *105*(8), 2644–78. https://doi.org/10.1257/aer.20110306

Coppock, A. (2023). *Persuasion in parallel*. University of Chicago Press.

Corbacho, A., Gingerich, D. W., Oliveros, V., & Ruiz-Vega, M. (2016). Corruption as a self-fulfilling prophecy: Evidence from a survey experiment in costa rica. *American Journal of Political Science*, *60*(4), 1077–1092. https://doi.org/10.1111/ajps.12244

Cornejo, R. C. (2022). Same scandal, different interpretations: Politics of corruption, anger, and partisan bias in mexico. *Journal of Elections, Public Opinion and Parties*, *0*(0), 1–22.

Coutts, A. (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, *22*(2), 369–395. https://doi.org/10.1007/s10683-018-9572-5

Cubel, M., Papadopoulou, A., & Sánchez-Pagés, S. (2024). Identity and political corruption: A laboratory experiment. *Economic Theory*.

D'Acunto, F., & Weber, M. (2024). Why survey-based subjective expectations are meaningful and important. *Annual Review of Economics*, *16*(Volume 16, 2024), 329–357. https://doi.org/https://doi.org/10.1146/annurev-economics-091523-043659

de Figueiredo, M. F., Hidalgo, F. D., & Kasahara, Y. (2022). When do voters punish corrupt politicians? experimental evidence from a field and survey experiment. *British Journal of Political Science*, 1–12. https://doi.org/10.1017/S0007123421000727

de Sousa, L., Clemente, F., & Maciel, G. G. (2023). Mapping conceptualisations and evaluations of corruption through survey questions: Five decades of public opinion-centred research. *European Political Science*, *22*(3), 368–383. https://doi.org/10.1057/s41304-023-00422-z

DellaVigna, S., & Gentzkow, M. (2010). Persuasion: Empirical evidence. *Annual Review of Economics*, *2*(1), 643–669.

Denly, M. (2022). Measuring corruption using governmental audits: A new framework and dataset [Accessed = 2022-11-01]. https://mikedenly.com/research/audit-measurement

Duch, R. M., Loewen, P., Robinson, T. S., & Zakharov, A. (2025). Governing in the face of a global crisis: When do voters punish and reward incumbent governments? *Proceedings of the National Academy of Sciences*, *122*(4), e2405021122. https://doi.org/10.1073/pnas.2405021122

Dunning, T., Grossman, G., Humphreys, M., Hyde, S., McIntosh, C., & Nellis, G. (2019). *Information, accountability, and cumulative learning: Lessons from metaketa i*. Cambridge University Press.

Dunning, T., Grossman, G., Humphreys, M., Hyde, S. D., McIntosh, C., Nellis, G., Adida, C. L., Arias, E., Bicalho, C., Boas, T. C., Buntaine, M. T., Chauchard, S., Chowdhury, A., Gottlieb, J., Hidalgo, F. D., Holmlund, M., Jablonski, R., Kramon, E., Larreguy, H., … Sircar, N. (2019). Voter information campaigns and political accountability: Cumulative findings from a preregistered meta-analysis of coordinated trials. *Science Advances*, *5*(7).

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, *3*(2), 114–38. https://doi.org/10.1257/mic.3.2.114

Elia, E., & Schwindt-Bayer, L. A. (2022). Corruption perceptions, opposition parties, and reelecting incumbents in latin america. *Electoral Studies*, *80*, 102545.

Enríquez, J. R., Larreguy, H., Marshall, J., & Simpser, A. (2024). Mass political information on social media: Facebook ads, electorate saturation, and electoral accountability in mexico. *Journal of the European Economic Association*, *22*(4), 1678–1722. https://doi.org/10.1093/jeea/jvae011

Erlich, A., Gans-Morse, J., & Nichter, S. (2025). Selective bribery: When do citizens engage in corruption? *Comparative Political Studies*, *58*(5), 996–1036. https://doi.org/10.1177/00104140241259444

Esberg, J., & Mummolo, J. (2018). *Explaining misperceptions of crime* [Unpublished manuscript].

Fan, T. Q., Liang, Y., & Peng, C. a. (2024). *The inference-forecast gap in belief updating* [Revise and resubmit at *Econometrica*. Latest draft and experimental instructions available online]. https://www.dropbox.com/scl/fi/7gyxuu80maruqbdqjo3lv/ifgap.pdf

Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice*, *50*, 5–25.

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

Ferraz, C., & Finan, F. (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review*, *101*(4), 1274–1311. https://doi.org/10.1257/aer.101.4.1274

Fiorina, M. P. (1981). *Retrospective voting in american national elections*. Yale University Press.

Gagliarducci, S., & Manacorda, M. (2020). Politics in the family: Nepotism and the hiring decisions of italian firms. *American Economic Journal: Applied Economics*, *12*(2), 67–95. https://doi.org/10.1257/app.20170778

Gerber, A. S., & Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company, Inc.

Ghanem, D., Hirshleifer, S., & Ortiz-Beccera, K. (2023). Testing attrition bias in field experiments. *Journal of Human Resources*. https://doi.org/10.3368/jhr.0920-11190R2

Gill, J., & Walker, L. D. (2005). Elicited priors for bayesian model specifications in political science research. *Journal of Politics*, *67*(3), 841–872. https://doi.org/https://doi.org/10.1111/j.1468-2508.2005.00342.x

Gillen, B., Snowberg, E., & Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, *127*(4), 1826–1863.

Gonzalez-Ocantos, E., de Jonge, C. K., & Nickerson, D. W. (2014). The conditionality of vote buying norms: Experimental evidence from latin america. *American Journal of Political Science*, *58*(1), 197–211.

Greifer, N. (2022). *Cobalt: Covariate balance tables and plots. r package version 4.4.1* (tech. rep.).

Haaland, I., & Roth, C. (2021). Beliefs about racial discrimination and support for pro-black policies. *The Review of Economics and Statistics*, 1–38. https://doi.org/10.1162/rest_a_01036

Haaland, I., Roth, C., & Wohlfart, J. (2023). Designing information provision experiments. *Journal of Economic Literature*, *61*(1), 3–40. https://doi.org/10.1257/jel.20211658

Healy, A., & Lenz, G. S. (2014). Substituting the end for the whole: Why voters respond primarily to the election-year economy. *American Journal of Political Science*, *58*(1), 31–47. https://doi.org/https://doi.org/10.1111/ajps.12053

Hicken, A., Leider, S., Ravanilla, N., & Yang, D. (2015). Measuring vote-selling: Field evidence from the philippines. *American Economic Review*, *105*(5), 352–56. https://doi.org/10.1257/aer.p20151033

Hopkins, D. J., Sides, J., & Citrin, J. (2019). The muted consequences of correct information about immigration. *The Journal of Politics*, *81*(1), 315–320.

Jablonski, R. S., Buntaine, M. T., Nielson, D. L., & Pickering, P. M. (2021). Individualized text messages about public services fail to sway voters: Evidence from a field experiment on ugandan elections. *Journal of Experimental Political Science*, 1–13.

Jäger, S., Roth, C., Roussille, N., & Schoefer, B. (2024). Worker beliefs about outside options*. *The Quarterly Journal of Economics*, *139*(3), 1505–1556. https://doi.org/10.1093/qje/qjae001

Jahnke, B., & Weisser, R. A. (2019). How does petty corruption affect tax morale in sub-saharan africa? *European Journal of Political Economy*, *60*, 101751.

Kayser, M. A., & Peress, M. (2012). Benchmarking across borders: Electoral accountability and the necessity of comparison. *American Political Science Review*, *106*(3), 661–684. https://doi.org/10.1017/S0003055412000275

Kovach, M. (2021, January). *Conservative Updating* (Papers No. 2102.00152). arXiv.org. https://ideas.repec.org/p/arx/papers/2102.00152.html

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Kuziemko, I., Norton, M. I., Saez, E., & Stantcheva, S. (2015). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review*, *105*(4), 1478–1508. https://doi.org/10.1257/aer.20130360

Lagunes, P. (2021, September). *The Eye and the Whip: Corruption Control in the Americas*. Oxford University Press. https://doi.org/10.1093/oso/9780197577622.001.0001

LAPOP. (2021). Chile: Americasbarometer 2021 [Vanderbilt University, Nashville, TN].

Larreguy, H., Marshall, J., & Snyder, J., James M. (2020). Publicising Malfeasance: When the Local Media Structure Facilitates Electoral Accountability in Mexico. *The Economic Journal*, *130*(631), 2291–2327. https://doi.org/10.1093/ej/ueaa046

Larsen, M. V., & Olsen, A. L. (2020). Reducing bias in citizens' perception of crime rates: Evidence from a field experiment on burglary prevalence. *The Journal of Politics*, *82*(2), 747–752. https://doi.org/10.1086/706595

Letki, N., Górecki, M. A., & Gendźwiłł, A. (2023). 'they accept bribes; we accept bribery': Conditional effects of corrupt encounters on the evaluation of public institutions. *British Journal of Political Science*, *53*(2), 690–697. https://doi.org/10.1017/S0007123422000047

Liang, Y. (2025). Learning from unknown information sources. *Management Science*, *71*(5), 3873–3890. https://doi.org/10.1287/mnsc.2021.03551

Martinelli, C. (2022). Accountability and grand corruption. *American Economic Journal: Microeconomics*, *14*(4), 645–79. https://doi.org/10.1257/mic.20200186

Mian, A., Sufi, A., & Khoshkhou, N. (2021). Partisan Bias, Economic Expectations, and Household Spending. *The Review of Economics and Statistics*, 1–46. https://doi.org/10.1162/rest_a_01056

Mutz, D. C. (2021). *Winners and losers: The psychology of foreign trade* (Vol. 27). Princeton University Press. https://doi.org/10.2307/j.ctv1fj85hw

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

OECD. (2016). *Supreme audit institutions and good governance*.

para la Transparencia, C. N. (2018). Estudio nacional de transparencia.

para la Transparencia, C. N. (2019). Estudio nacional de transparencia.

para la Transparencia, C. N. (2020). Estudio nacional de transparencia.

Peiffer, C. (2020). Message received? experimental findings on how messages about corruption shape perceptions. *British Journal of Political Science*, *50*(3), 1207–1215. https://doi.org/10.1017/S0007123418000108

Peterson, E., & Iyengar, S. (2021). Partisan gaps in political information and information-seeking behavior: Motivated reasoning or cheerleading? *American Journal of Political Science*, *65*(1), 133–147.

Peyton, K. (2020). Does trust in government increase support for redistribution? evidence from randomized survey experiments. *American Political Science Review*.

Rathje, S., Roozenbeek, J., Bavel, J. J. V., & van der Linden, S. (2023). Accuracy and social motivations shape judgements of (mis)information. *Nature Human Behaviour*, *7*(6), 892–903. https://doi.org/10.1038/s41562-023-01540-w

Reinikka, R., & Svensson, J. (2006). Using micro-surveys to measure and explain corruption [Part Special Issue (pp. 324–404). Corruption and Development: Analysis and Measurement]. *World Development*, *34*(2), 359–370. https://doi.org/https://doi.org/10.1016/j.worlddev.2005.03.009

Ridgeway, G., McCaffrey, D., Morral, A., Cefalu, M., Burgette, L., Pane, J., & Griffin, B. A. (2021, October). *Toolkit for weighting and analysis of nonequivalent groups: A guide to the twang package* (R Package). RAND. https://cran.r-project.org/web/packages/twang/vignettes/twang.pdf

Rodríguez, I., Rodon, T., Unan, A., Herbig, L., Klüver, H., & Kuhn, T. (2025). Benchmarking pandemic response: How the uk's covid-19 vaccine rollout impacted diffuse and specific support for the eu. *British Journal of Political Science*, *55*, e35. https://doi.org/10.1017/S0007123424000802

Rodríguez Chatruc, M., Stein, E., & Vlaicu, R. (2021). How issue framing shapes trade attitudes: Evidence from a multi-country survey experiment. *Journal of International Economics*, *129*, 103428. https://doi.org/https://doi.org/10.1016/j.jinteco.2021.103428

Roth, C., Settele, S., & Wohlfart, J. (2022). Beliefs about public debt and the demand for government spending [Annals Issue: Subjective Expectations  Probabilities in Economics]. *Journal of Econometrics*, *231*(1), 165–187. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.09.011

Roth, C., & Wohlfart, J. (2020). How do expectations about the macroeconomy affect personal expectations and behavior? *Review of Economics and Statistics*, *102*(4), 731–748. https://doi.org/10.1162/rest_a_00867

Sanna, G. A., & Lagnado, D. (2025). Belief updating in the face of misinformation: The role of source reliability. *Cognition*, *258*, 106090. https://doi.org/https://doi.org/10.1016/j.cognition.2025.106090

Solaz, H., Vries, C. E. D., & de Geus, R. A. (2019). In-group loyalty and the punishment of corruption. *Comparative Political Studies*, *52*(6), 896–926.

Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology*, *66*(8), S84–S90.e1.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755–769. https://doi.org/https://doi.org/10.1111/j.1540-5907.2006.00214.x

Tappin, B. M., Pennycook, G., & Rand, D. G. (2020). Bayesian or biased? analytic thinking and political belief updating [Epub 2020 Jun 24]. *Cognition*, *204*, 104375. https://doi.org/10.1016/j.cognition.2020.104375

Thaler, M. (2024). The fake news effect: Experimentally identifying motivated reasoning using trust in news. *American Economic Journal: Microeconomics*, *16*(2), 1–38. https://doi.org/10.1257/mic.20220146

Transparency International. (2023). *Corruption perceptions index 2022* (Accessed: 14 May 2025). Transparency International. https://images.transparencycdn.org/images/Report_CPI2022_English.pdf

World Bank. (2021, July). Supreme audit institutions independence index: 2021 global synthesis report.

Zappalà, G. (2024). Adapting to climate change accounting for individual beliefs. *Journal of Development Economics*, *169*, 103289. https://doi.org/https://doi.org/10.1016/j.jdeveco.2024.103289

Zhou, F., & Oostendorp, R. (2014). Measuring true sales and underreporting with matched firm-level survey and tax office data. *The Review of Economics and Statistics*, *96*(3), 563–576. https://doi.org/10.1162/REST_a_00408

Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review*, *110*(2), 337–61. https://doi.org/10.1257/aer.20180728

[letterpaper,12pt,leqno]article paper,appendix times xr lscape float setspace  paper [style=apa, backend=biber]biblatex manuscript.bib tabularray float graphicx codehigh [normalem]ulem

# Government Audits of Municipal Corruption and Belief Updating: Experimental Evidence.

Felipe Torres-Raposo

London School of Economics and Political Science

Raymond Duch

Nuffield College

November 28, 2025

## Contents

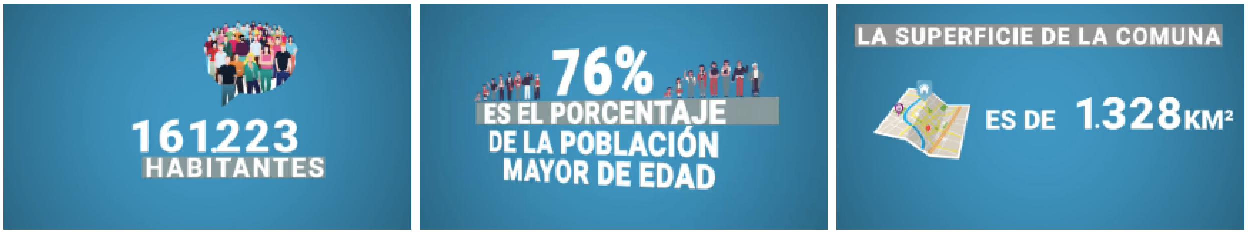
*f.torres-raposo@lse.ac.uk
f.torres-raposo@lse.ac.uk

# A Belief Updating

## A.1 Placebo Screen Shots

Figure A.1 shows some screenshots of the placebo treatment. The video provided three pieces of information: 1) The number of people that live in the commune, 2) the percentage of people above the legal age, and the size of the commune in squared meters.

Figure A.1: Screen Shoots - Placebo Treatment



*Note:* This figure shows example screenshots of the placebo treatment arm. Subjects in the placebo condition received three pieces of information about their municipality. The first screenshot (left) reported the population that resides in the municipality. The second screenshot (middle) reported the population that is above the legal age (18 years old). The third (right) and final piece of information is the size of the commune in terms of squared kilometers.

## A.2 Subject Recruitment

The initial phase of the experiment took place from late September 2019 to January 2020, during which we recruited 49,883 Chilean residents aged 18 or over as potential participants. We recruited subjects using Facebook, Instagram, and Twitter ads. We complemented our online recruitment efforts with traditional forms of subject recruitment, including fliers, posters, and newspaper advertisements. Study participants completed a brief form that requested information about the municipality where they reside, their gender, and contact details. In January 2020, the CGO did a final section of the 116 municipalities that would be audited in 2020. Our recruitment efforts focused on these communes.

We recruited our sample from a population of Chileans who have a social media presence, particularly on Facebook. The recruitment took place over two phases. In the first phase, we

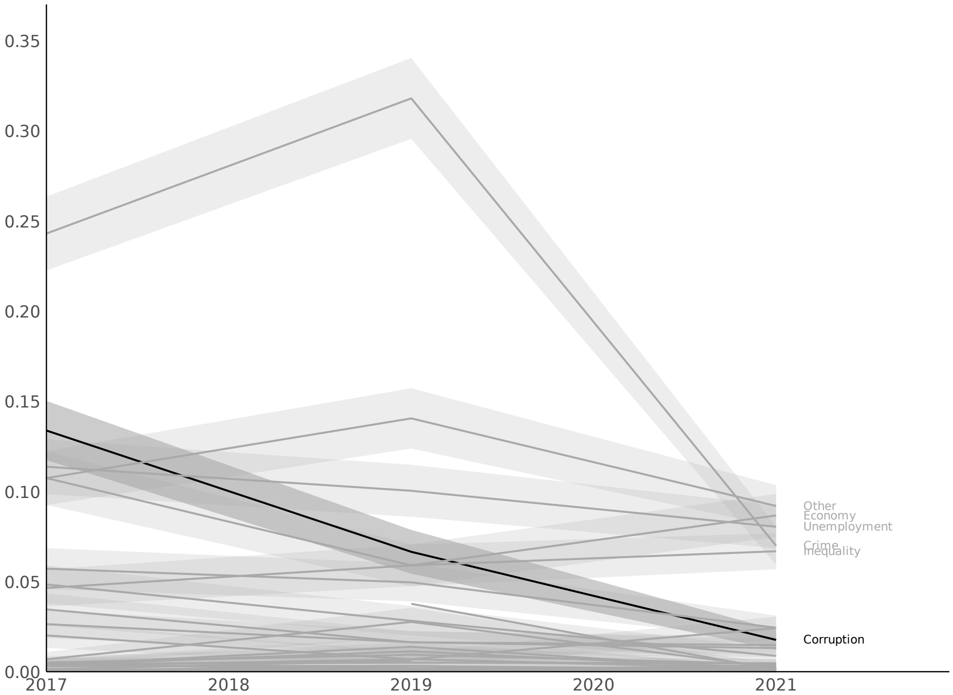
Table A.1: Profile of Final Sample

| | Census | CASEN | Sample |
|---|---|---|---|
| N | 17,574,003 | 2,164,439 | 5,528 |
| Age | 40 | 37 | 36 |
| Education | | | |
| Secondary | 40% | 37% | 67% |
| Tertiary | 24% | 22% | 31% |
| Female | 51% | 52% | 65% |
| Income (USD) | - | 607 | 658 |

*Note:* In this table, we compare our sample to Chile's population, examining key demographics such as age, income, educational attainment, and gender. The first column provides population estimates of these covariates using data from the 2017 census. The second column includes population estimates based on the 2020 National Socio-Economic Characterization Survey. Age is the age of respondents in 2020. Education is a categorical variable with levels: ”Primary,” ”Secondary,” and ”Tertiary.” We only report ”Secondary” and ”Tertiary.” ”Primary” is the reference category. Income is the respondents' monthly income reported in USD dollars (Exchange rate  771 Chilean pesos equal to 1 USD).

recruited 49,883 subjects into a pool of potential subjects. A total of 22,755 of these recruits resided in the 116 municipalities audited by the CGO and, hence, were eligible for the experiment. We sent invitations to participate in the pre-treatment survey to these 22,755 eligible subjects. We achieved a response rate of 25%, resulting in 5,528 subjects participating in the pre-treatment survey. Table A.1 compares the demographic features of the final sample with statistics from the Chilean national census and the National Socio-Economical Characterization Survey. Note that the pool and the sample are very similar in terms of age, education, gender, and income. There are differences between our sample and Chile's demographic population. Our sample is disproportionately younger and has higher educational attainment. In terms of gender, the study had a higher proportion of women, representing 65% of our sample. Our sample reports have a slightly higher average income income compared to the population's average.

Figure A.2: Saliency of Corruption in Chile - AmericasBarometer Surveys

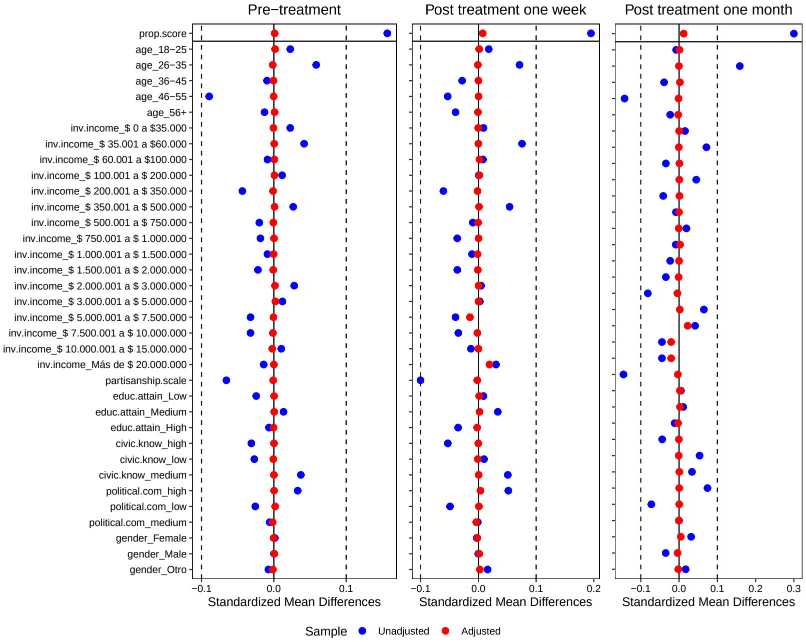


*Note:* This graph displays the proportion of respondents who identified the most pressing issue facing the country. The data come from the 2017, 2019, and 2021 waves of the AmericasBarometer, conducted by the Latin American Public Opinion Project (LAPOP, 2021). No survey was conducted in 2020. Respondents were asked the following question: *In your opinion, what is the most serious problem the country is currently facing?* Estimates are weighted using the sampling weights provided in the data to attain national representative estimates.

## A.3 Treatment Random Assignment and Attrition

**Balance** Respondents are observed at three different time points during the trial: baseline, when treatment is assigned and implemented; one week post-treatment; and one month post-treatment. Balance on covariates is assessed in Figure A.3 by comparing (in red) standardized mean differences (raw differences in proportion for binary variables) for the treatment and placebo arms. The estimates are generated by the R package *cobalt* (Greifer, 2022). We also compare these unadjusted differences with those obtained when the sample is weighted using propensity score matching (in green). These comparisons are generated for the sample of 6,050 participants interviewed at baseline, the 3,592 participants interviewed one-week post-treatment, and the 2,255 participants contacted one-month post-treatment. We employ indicative balance tolerance levels of 0.1 (Stuart et al., 2013) in Figure A.3 (the vertical dotted lines). With only a couple of exceptions, the unadjusted sample standardized mean differences for covariates across the three comparisons fall within this 0.1 threshold. The exceptions are the partisanship scale and the 25-35 age category, both of which are slightly lower than the -0.1 threshold in the one-month post-treatment follow-up sample.

Figure A.3: Balance on Undjusted and Adjusted Standardized Mean Differences



*Note:* This figure shows unadjusted (blue) and adjusted (red) standardized mean differences for key covariates. Adjusted values are estimated using nearest neighbor matching based on propensity scores from a logistic regression predicting treatment. Covariates include age, gender, income, partisanship, education, civic, and political knowledge

Table A.2 reports the results of regressing subjects' treatment assignment (treatment versus placebo) on observed covariates, including both nested and full models. The initially assigned sample of subjects is balanced – only one of the covariate predictor variables, partisanship, is weakly significant. Table A.2 also indicates that with rising attrition and declining compliance, the resulting samples remain balanced across the two treatment groups in most covariates. We find differences between the two groups for *Civic knowledge* and *Partisanship* covariates in the follow-up survey. This result indicates that respondents who received the treatment had lower levels of civic knowledge compared to the placebo group and lower partisanship.

Table A.2: Randomization Checks - Pre-treatment Variables

| | Baseline | Post-Treatment | Follow-up |
|---|---|---|---|
| Intercept | 0.661*** | 0.692*** | 0.628*** |
| | (0.022) | (0.035) | (0.054) |
| Gender | 0.007 | 0.005 | -0.018 |
| | (0.013) | (0.017) | (0.022) |
| Income | -0.005 | -0.009 | -0.009 |
| | (0.014) | (0.018) | (0.031) |
| Education | 0.003 | -0.007 | -0.010 |
| | (0.015) | (0.022) | (0.029) |
| Partisanship | -0.006** | -0.010** | -0.015** |
| | (0.003) | (0.004) | (0.006) |
| Turnout | -0.014 | -0.016 | -0.001 |
| | (0.013) | (0.019) | (0.023) |
| Support incumbent | -0.011 | -0.011 | -0.010 |
| | (0.014) | (0.021) | (0.036) |
| Civic knowledge: Medium | 0.038 | -0.040 | -0.160* |
| | (0.058) | (0.073) | (0.088) |
| Civic knowledge: High | 0.027 | -0.066 | -0.184* |
| | (0.058) | (0.075) | (0.090) |
| Political information: Medium | 0.003 | 0.009 | 0.019 |
| | (0.013) | (0.015) | (0.021) |
| Political information: High | 0.016 | 0.032 | 0.046 |
| | (0.016) | (0.023) | (0.032) |
| Anova F-test Nested (P-value) | 0.353 | 0.07 | 0.045* |
| Num.Obs. | 6050 | 3592 | 2255 |
| R2 | 0.002 | 0.004 | 0.008 |
| R2 Adj. | 0.000 | 0.001 | 0.003 |

*Note:* This table reports regression estimates of treatment assignment on pre-treatment variables. Treatment was assigned to 6,050 respondents, but 525 were excluded due to incomplete audits. Standard errors are clustered by municipality. The table also shows p-values from F-tests comparing models with and without covariates. $*p < .05$, $**p < .01$, $*p < .001$

**Attrition** As we point out in the main text, we observe attrition in both treatment and placebo conditions. In Table A.4, we report pre-treatment measures of respondents' outcomes and covariates for the whole, observed, and attriters samples across all waves. This analysis shows near-zero differences for all three outcomes between attriters and the observed sample in post-treatment and follow-up waves. Similarly, we examined individual-level covariates, including gender, education, income, and co-partisanship. We did not observe statistically significant differences across all covariates. There is evidence that higher-income subjects were more likely to drop out in the post-treatment and follow-up surveys.

We further examined differential attrition by treatment status. Table A.5 compares pre-treatment outcomes and covariates across treatment and placebo groups, distinguishing between attriters and observed respondents. We find no systematic baseline differences, except for the Qualitative outcome in the treatment group, where attriters reported higher prior corruption beliefs. Pre-treatment beliefs for the two quantitative outcomes remain balanced across all groups.

We find no differences in behavioral outcomes between attriters and non-attriters within treatment arms across both waves. At the individual level, groups are similar in terms of gender, education, co-partisanship, and income. However, among treated respondents, attriters in the post-treatment survey tend to be older and more right-leaning. As noted, there is some evidence that attrition rates are higher among treated respondents living in high-corruption communities.

We assessed whether key covariates and treatment assignment were correlated with differential attrition. Table A.3 presents logistic regression results, where the outcome is a binary indicator for post-treatment survey dropout. Attrition is positively associated with treatment assignment and baseline donations but not with other covariates such as co-partisanship, age, education, or political knowledge. Table A.6 reports complementary tests by block design.

A potential threat to internal validity is that attrition may be correlated with treatment assignment. To mitigate this, we block-randomized respondents based on prior beliefs about municipal malfeasance, forming "Low" and "High" blocks within each commune. Table A.6 shows no evidence of differential attrition across these blocks. We also examined whether attrition was associated with block size, which varied due to our randomization design. Results in columns 2 and 3 show that neither block type nor size predicts attrition. Finally, we tested whether attrition was related to recruitment intensity, measured by variation in sample sizes across the 116 communes. Column 4 shows no association, suggesting that attrition is not systematically related to recruitment procedures.

Table A.3: Attrition Covariates

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| (Intercept) | -1.010*** | -1.329*** | -1.994*** | -1.784** | -0.956 |
| | (0.053) | (0.349) | (0.501) | (0.572) | (0.844) |
| Treat | 0.569*** | 0.552*** | 0.593*** | 0.593*** | -0.755 |
| | (0.063) | (0.064) | (0.067) | (0.067) | (0.955) |
| Age (Years) | | -0.002 | -0.003 | -0.002 | -0.004 |
| | | (0.003) | (0.003) | (0.003) | (0.005) |
| Log Income | | 0.066** | 0.072** | 0.072** | 0.109* |
| | | (0.022) | (0.023) | (0.023) | (0.044) |
| Partisanship | | 0.005 | -0.007 | -0.005 | -0.030 |
| | | (0.015) | (0.016) | (0.016) | (0.028) |
| Education:Medium | | -0.448 | -0.396 | -0.388 | -0.810 |
| | | (0.258) | (0.272) | (0.272) | (0.446) |
| Education:High | | -0.421 | -0.355 | -0.347 | -0.965* |
| | | (0.253) | (0.267) | (0.267) | (0.437) |
| Co-Partisan | | 0.038 | 0.015 | 0.011 | -0.001 |
| | | (0.065) | (0.073) | (0.073) | (0.128) |
| Civic knowledge:Medium | | | | -0.299 | -0.912* |
| | | | | (0.272) | (0.428) |
| Civic knowledge:High | | | | -0.372 | -1.053* |
| | | | | (0.273) | (0.428) |
| Political knowledge:Medium | | | | 0.096 | 0.256 |
| | | | | (0.075) | (0.142) |
| Political knowledge:High | | | | 0.112 | 0.327 |
| | | | | (0.089) | (0.168) |
| Treat:Age(Years) | | | | | 0.003 |
| | | | | | (0.006) |
| Treat:Log Income | | | | | -0.051 |
| | | | | | (0.052) |
| Treat:Partisanship | | | | | 0.037 |
| | | | | | (0.034) |
| Treat:Education:Medium | | | | | 0.671 |
| | | | | | (0.562) |
| Treat:Education:High | | | | | 0.944 |
| | | | | | (0.552) |
| Treat:Partisan | | | | | 0.022 |
| | | | | | (0.150) |
| Treat:Civic knowledge:Medium | | | | | 0.968 |
| | | | | | (0.554) |
| Treat:Civic knowledge:High | | | | | 1.060 |
| | | | | | (0.555) |
| Treat:Political Knowledge:Medium | | | | | -0.221 |
| | | | | | (0.168) |
| Treat:Political Knowledge:High | | | | | -0.301 |
| | | | | | (0.198) |
| Commune FE | No | No | Yes | Yes | Yes |
| Num.Obs. | 5528 | 5310 | 5310 | 5310 | 5310 |
| AIC | 7078.1 | 6816.8 | 6710.3 | 6713.7 | 6720.0 |
| BIC | 7091.3 | 6869.4 | 7519.3 | 7549.0 | 7621.1 |

*Note:* This table reports the results of three logistic regressions, where the dependent variable is a binary variable equal to 1 if the respondent is an attritor and 0 if not, in the post-treatment survey. The independent variable is the treatment assignment variable *Treat*. Column 1 represents the baseline model, which includes only the *Treat* variable. Column 2 summarizes the regression results after adding additional covariates, such as age, co-partisanship, and education. The model in column 3 expands from the model in column 2, controlling for *commune* fixed effects. The model in column 4 builds upon model 3, incorporating additional controls, such as civic and political knowledge measures. Finally, in column 5, we estimated a fully saturated model that interacts with the treatment variable with all covariates. $*p < .05, **p < .01, ***p < 0.001$.

Table A.4: Pre-Treatment Descriptive Analysis for Malfeasance Belief Outcome Measures and Covariates for observed Samples and Attriters

| | Baseline | Post-Treat | | | Follow-up | | |
|---|---|---|---|---|---|---|---|
| | Whole sample | Observed sample | Attriters | Difference | Observed sample | Attriters | Difference |
| **Qualitative Measures** | | | | | | | |
| Subjective malfeasance scale 1-10 | 4.49 | 4.45 | 4.55 | 0.10 | 4.51 | 4.47 | -0.05 |
| | (2.44) | (2.40) | (2.51) | | (2.40) | (2.47) | |
| Certainty Subjective malfeasance scale 1-10 | 5.40 | 5.38 | 5.45 | 0.07 | 5.40 | 5.41 | 0.00 |
| | (2.71) | (2.70) | (2.71) | | (2.70) | (2.71) | |
| **Quantitative Measures - Incentivized** | | | | | | | |
| Resources (1-10) | 3.84 | 3.85 | 3.84 | -0.01 | 3.85 | 3.84 | -0.01 |
| | (2.38) | (2.40) | (2.35) | | (2.38) | (2.38) | |
| Distribution (1-4) | 1.90 | 1.89 | 1.93 | 0.04 | 1.90 | 1.91 | 0.01 |
| | (0.78) | (0.79) | (0.78) | | (0.78) | (0.79) | |
| **Behavioral measures** | | | | | | | |
| 500 CLP Municipal Donation | 32% | 31% | 34% | 3% | 31% | 33% | 3% |
| | (0.63%) | (1.08%) | (0.77%) | | (0.83%) | ( 0.01%) | |
| **Covariates** | | | | | | | |
| Female | 64.73 | 64.67 | 64.82 | 0.15 | 65.76 | 64.01 | -1.75 |
| | (0.64) | (1.09) | (0.80) | | (1.00) | (0.84) | |
| Education (Yrs) | 15.07 | 15.06 | 15.07 | 0.01 | 15.05 | 15.08 | 0.03 |
| | (0.04) | (0.07) | (0.05) | | (0.05) | (0.06) | |
| Age | 35.88 | 35.76 | 36.11 | 0.35 | 35.77 | 35.96 | 0.19 |
| | (0.18) | (0.29) | (0.22) | | (0.23) | (0.28) | |
| Partisanship (1-10) | 4.47 | 4.46 | 4.49 | 0.03 | 4.43 | 4.49 | 0.06 |
| | (0.03) | (0.05) | (0.03) | | (0.04) | (0.04) | |
| Support for the incumbent | 27.13% | 27.00% | 27.38% | 0.38% | 26.87% | 27.31% | 0.44% |
| | (0.59%) | (1.01%) | (0.74%) | | (0.78%) | (0.93%) | |
| Income (USD) | 658 | 896 | 1012 | 116 | 775 | 1046 | 271 |
| | (70) | (77) | (140) | | (18) | (117) | |
| **Municipal-level covariates** | | | | | | | |
| Share previous malfeasance | 0.04 | 0.03 | 0.04 | 0.01 | 0.03 | 0.04 | 0.01 |
| | (0.14) | (0.13) | (0.15) | | (0.11) | (0.16) | |
| Historical irregularities | 43.03 | 42.47 | 44.08 | 1.61 | 42.56 | 43.35 | 0.79 |
| | (16.51) | (16.64) | (17.31) | | (16.51) | (17.15) | |
| Sample | 5528 | 3592 | 1936 | | 2255 | 3273 | |

*Note:* This table reports pre-treatment values for all outcomes and relevant covariates broken down by attrition status. Along with mean values, we also calculate the difference in means for each variable. Standard errors are reported in parenthesis.

Table A.5: Descriptive for Malfeasance Belief Outcome Measures and Covariates for Samples and Attriters by Treatment Status

| | | | Post-Treat | | | | | | Follow-up | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Baseline | | Observed Sample | | Attriters | | Differences | | Observed Sample | | Attriters | | Differences | |
| | Placebo | Treatment | Placebo | Treatment | Placebo | Treatment | Placebo | Treat | Placebo | Treat | Placebo | Treatment | Placebo | Treat |
| **Qualitative Measures** | | | | | | | | | | | | | | |
| Subjective malfeasance scale (1-10) | 4.47 | 4.49 | 4.50 | 4.42 | 4.42 | 4.60 | 0.08 | -0.18 | 4.53 | 4.50 | 4.40 | 4.49 | 0.13 | 0.01 |
| | (2.44) | (2.44) | (2.39) | (2.41) | (2.56) | (2.49) | | | (2.43) | (2.37) | (2.45) | (2.48) | | |
| Certainty of Subjective malfeasance scale (1-10) | 5.44 | 5.39 | 5.45 | 5.34 | 5.41 | 5.46 | 0.05 | -0.12 | 5.52 | 5.30 | 5.33 | 5.43 | 0.19 | -0.13 |
| | (2.72) | (2.70) | (2.73) | (2.69) | (2.70) | (2.72) | | | (2.73) | (2.66) | (2.70) | (2.72) | | |
| **Quantitative Measures - Incentivized** | | | | | | | | | | | | | | |
| Resources (1-10) | 3.83 | 3.85 | 3.84 | 3.85 | 3.83 | 3.85 | 0.01 | 0.01 | 3.82 | 3.88 | 3.86 | 3.83 | -0.04 | 0.05 |
| | (2.36) | (2.39) | (2.35) | (2.42) | (2.39) | (2.33) | | | (2.33) | (2.42) | (2.40) | (2.37) | | |
| Distribution (1-4) | 1.90 | 1.91 | 1.90 | 1.89 | 1.91 | 1.94 | -0.01 | -0.05 | 1.91 | 1.90 | 1.90 | 1.91 | 0.01 | -0.01 |
| | (0.78) | (0.79) | (0.78) | (0.79) | (0.79) | (0.78) | | | (0.78) | (0.77) | (0.78) | (0.79) | | |
| **Behavioral measures** | | | | | | | | | | | | | | |
| 500 CLP Municipal Donation | 32% | 33% | 32% | 31% | 33% | 35% | -1% | -4% | 31% | 31% | 33% | 33% | -2% | -2% |
| | (1.09%) | (0.77%) | (2.12%) | (1.26%) | (1.27%) | (0.98%) | | | (1.69%) | (0.94%) | (1.42%) | (1.34%) | | |
| **Covariates** | | | | | | | | | | | | | | |
| Female | 64.79 | 64.69 | 64.78 | 64.61 | 64.83 | 64.82 | -0.05 | -0.22 | 64.81 | 66.61 | 64.77 | 63.77 | 0.03 | 2.84 |
| | (1.12%) | (0.79%) | (2.16%) | (1.26%) | (1.30%) | (1.01%) | | | (1.72%) | (0.96%) | (1.47%) | (1.36%) | | |
| Education (Yrs) | 15.09 | 15.05 | 15.10 | 15.04 | 15.06 | 15.07 | 0.04 | -0.03 | 15.06 | 15.03 | 15.13 | 15.06 | -0.06 | -0.03 |
| | (0.07) | (0.05) | (0.13) | (0.08) | (0.08) | (0.06) | | | (0.1) | (0.06) | (0.08) | (0.08) | | |
| Age | 36.42 | 35.61 | 36.41 | 35.37 | 36.46 | 35.99 | -0.05 | -0.62 | 36.45 | 35.17 | 36.38 | 35.83 | 0.07 | -0.66 |
| | (0.07) | (0.05) | (0.57) | (0.34) | (0.37) | (0.28) | | | (0.46) | (0.26) | (0.43) | (0.38) | | |
| Partisanship | 4.55 | 4.43 | 4.58 | 4.38 | 4.47 | 4.50 | 0.11 | -0.12 | 4.58 | 4.31 | 4.51 | 4.49 | 0.07 | -0.18 |
| | (0.05) | (0.03) | (0.09) | (0.05) | (0.06) | (0.04) | | | (0.07) | (0.04) | (0.06) | (0.06) | | |
| Co-partisanship | 27.78% | 26.81% | 27.92% | 26.46% | 27.40% | 27.37% | 0.52% | -0.91% | 27.63% | 26.21% | 28.00% | 27.10% | -0.37% | -0.89% |
| | (1.05%) | (0.73%) | (0.09%) | (0.05%) | (0.06%) | (0.04%) | | | (0.07%) | (0.04%) | (0.06%) | (0.06%) | | |
| Income (USD) | 1003 | 904 | 853 | 921 | 1404 | 878 | 551 | -43 | 793 | 973 | 1282 | 973 | 489 | 214 |
| | (149) | (74) | (544) | (25) | (28) | (121) | | | (346) | (108) | (26) | (24) | | |
| **Municipal-level covariates** | | | | | | | | | | | | | | |
| Share previous malfeasance | 0.04 | 0.03 | 0.04 | 0.03 | 0.05 | 0.04 | 0.01 | 0.01 | 0.04 | 0.02 | 0.04 | 0.04 | 0.00 | 0.02 |
| | (0.15) | (0.13) | (0.14) | (0.12) | (0.19) | (0.14) | | | (0.14) | (0.04) | (0.16) | (0.15) | | |
| Historical irregularities | 42.53 | 43.28 | 42.27 | 42.58 | 43.23 | 44.37 | 0.26 | 1.79 | 42.14 | 42.92 | 42.92 | 43.45 | 0.78 | 0.53 |
| | (17.1) | (16.8) | (17.04) | (16.41) | (17.2) | (17.34) | | | (16.82) | (16.23) | (17.43) | (17.06) | | |
| Sample | 1832 | 3696 | 1343 | 2249 | 489 | 2249 | | | 1057 | 1198 | 775 | 1198 | | |

*Note:* This table reports pre-treatment values for all outcomes and relevant covariates broken down by attrition status and treatment status. Along with mean values, we report standard deviations in parentheses. We also calculate the difference in means for each variable, grouped by their attrition and treatment status.

Table A.6: Attrition Test - Block type and Block Size and Commune Sample

| | Post-treatment | | | | Follow-up | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| (Intercept) | 0.345*** | 0.386*** | 0.383*** | 0.398*** | 0.591*** | 0.615*** | 0.628*** | 0.626*** |
| | (0.013) | (0.025) | (0.036) | (0.039) | (0.025) | (0.037) | (0.057) | (0.060) |
| Block - High | 0.019 | | 0.038 | | 0.002 | | -0.034 | |
| | (0.017) | | (0.042) | | (0.024) | | (0.071) | |
| Block size | | -0.001 | -0.001 | | | 0.000 | -0.001 | |
| | | (0.001) | (0.001) | | | (0.001) | (0.001) | |
| Block size:Block-High | | | -0.002 | | | | 0.001 | |
| | | | (0.001) | | | | (0.003) | |
| Sample commune | | | | -0.001 | | | | 0.000 |
| | | | | (0.001) | | | | (0.001) |
| Num.Obs. | 5528 | 5528 | 5528 | 5528 | 5528 | 5528 | 5528 | 5528 |
| R2 | 0.000 | 0.002 | 0.003 | 0.002 | 0.000 | 0.001 | 0.001 | 0.001 |
| R2 Adj. | 0.000 | 0.002 | 0.002 | 0.002 | 0.000 | 0.001 | 0.000 | 0.001 |

*Note:* This table reports the results of three models using OLS. The dependent variable is a binary variable, equal to 1 if the respondent is an attriter and 0 if not, in the post-treatment survey. The independent variable is the block randomization variable used to conduct complete randomization. Each commune had two blocks: *Low* and *High*, where high corresponded to the block that grouped respondents with low prior beliefs about malfeasance. At the same time, the high block grouped respondents with high prior beliefs. We also evaluated whether the block size predicts differential attrition. The independent variable in this model is the sample within each commune and each block. Finally, in model *Both*, we interacted with the block type (High or Low) and block size variables. Standard errors are clustered at the commune level. $*p < .05, **p < .01, ***p < 0.001$.

**Inverse Probability Weighting** With the complete set of demographic measures collected in the pre-treatment survey we can model attrition quite robustly. Inverse probability weighting is one of the estimation strategies we can implement. It has the value of being very straightforward – essentially modeling the attrition process as a function of observable covariates Anderson et al. (2021). The weights are based on the predicted values from a logistic regression of a binary variable indicating whether the observation is missing on the available covariates.[1] The weight is simply 1 over 1 minus these predicted probabilities.[2] We re-estimate treatment effects on the subset of the data where outcomes are observed and weight that estimate using these weights. The estimate from this regression is a consistent estimate for the treatment effect, assuming the censoring process is observable. Table A.7 compares the unweighted baseline treatment model

[1] An alternative here that we would consider using is a propensity score estimation algorithm to fully model any possible nonlinearities – the *twang* package, for example (Ridgeway et al., 2021).

[2] These weights are often characterized as “Weighted” – a slightly modified estimation strategy can generate more “stable” weights.

coefficients with those that incorporate IPW for the three outcome variables. The coefficients for the weighted estimates are very similar to those for the unweighted estimates. In particular, the estimated treatment effects are similar in both weighted and unweighted estimations.

Table A.7: Inverse Probability Weighting Model Regressions and Unweighted Models

| **Unweighted** | Information | Magnitudes | Negative | Positive | Information | Magnitudes | Negative | Positive | Information | Magnitudes | Negative | Positive |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | 2.734*** | 2.717*** | 2.826*** | 2.527+ | 2.302*** | 2.274*** | 2.333*** | 2.481 | 1.122*** | 1.125*** | 1.191*** | 0.861+ |
|  | (0.401) | (0.402) | (0.407) | (1.341) | (0.662) | (0.648) | (0.686) | (1.969) | (0.153) | (0.152) | (0.161) | (0.459) |
| Prior | 0.467*** | 0.464*** | 0.473*** | 0.436*** | 0.446*** | 0.444*** | 0.442*** | 0.458*** | 0.421*** | 0.415*** | 0.406*** | 0.453*** |
|  | (0.020) | (0.020) | (0.022) | (0.055) | (0.018) | (0.018) | (0.019) | (0.054) | (0.024) | (0.024) | (0.024) | (0.093) |
| Treat | 0.749*** | 0.736*** | 0.773*** | 0.566** | 0.740*** | 0.739*** | 0.834*** | 0.316 | 0.347*** | 0.327*** | 0.375*** | 0.073 |
|  | (0.079) | (0.091) | (0.103) | (0.203) | (0.088) | (0.094) | (0.103) | (0.200) | (0.028) | (0.035) | (0.036) | (0.081) |
| Stand - Malfeasance |  | 0.490 | 0.334 | -1.877 |  | 0.590 | 0.559 | 2.175 |  | 0.217 | 0.220 | -9.230+ |
|  |  | (0.489) | (0.476) | (22.915) |  | (0.432) | (0.448) | (16.192) |  | (0.233) | (0.240) | (5.108) |
| Treat:Stand - Malfeasance |  | 0.179 | 0.065 | 4.383 |  | 0.012 | -0.259 | 11.969 |  | 0.302 | 0.158 | 8.216 |
|  |  | (0.882) | (0.904) | (14.757) |  | (0.597) | (0.518) | (16.308) |  | (0.353) | (0.294) | (5.262) |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Num.Obs. | 3439 | 3439 | 2941 | 498 | 3439 | 3439 | 2941 | 498 | 3439 | 3439 | 2941 | 498 |
| R2 | 0.263 | 0.264 | 0.270 | 0.268 | 0.220 | 0.221 | 0.217 | 0.304 | 0.203 | 0.207 | 0.204 | 0.244 |
| R2 Adj. | 0.257 | 0.258 | 0.262 | 0.226 | 0.214 | 0.214 | 0.209 | 0.264 | 0.196 | 0.200 | 0.196 | 0.201 |
| **Weighted** |  |  |  |  |  |  |  |  |  |  |  |  |
| Intercept | 2.684*** | 2.667*** | 2.490*** | 2.311*** | 1.107*** | 1.999*** | 2.036*** | 1.776*** | 1.108*** | 1.127*** | 1.144*** | 1.086*** |
|  | (0.415) | (0.417) | (0.133) | (0.233) | (0.151) | (0.098) | (0.116) | (0.209) | (0.152) | (0.053) | (0.056) | (0.169) |
| Prior | 0.465*** | 0.463*** | 0.479*** | 0.446*** | 0.423*** | 0.457*** | 0.454*** | 0.477*** | 0.422*** | 0.422*** | 0.415*** | 0.459*** |
|  | (0.020) | (0.020) | (0.022) | (0.050) | (0.024) | (0.018) | (0.020) | (0.062) | (0.024) | (0.025) | (0.026) | (0.093) |
| Treat | 0.740*** | 0.731*** | 0.788*** | 0.577** | 0.346*** | 0.740*** | 0.824*** | 0.308 | 0.346*** | 0.328*** | 0.376*** | 0.066 |
|  | (0.079) | (0.090) | (0.105) | (0.209) | (0.028) | (0.094) | (0.104) | (0.192) | (0.028) | (0.036) | (0.037) | (0.075) |
| Stand - Malfeasance |  | 0.514 | 0.319 | 2.918 |  | 0.562 | 0.488 | 7.757 |  | 0.201 | 0.201 | -8.184 |
|  |  | (0.475) | (0.513) | (23.381) |  | (0.460) | (0.483) | (20.258) |  | (0.240) | (0.247) | (5.841) |
| Treat:Stand - Malfeasance |  | 0.133 | -0.083 | 0.414 |  | -0.111 | -0.366 | 9.287 |  | 0.300 | 0.154 | 8.092* |
|  |  | (0.846) | (0.856) | (12.716) |  | (0.626) | (0.562) | (18.324) |  | (0.358) | (0.298) | (3.496) |
| Num.Obs. | 3439 | 3439 | 2941 | 498 | 3439 | 3439 | 2941 | 498 | 3439 | 3439 | 2941 | 498 |
| R2 | 0.262 | 0.263 | 0.259 | 0.222 | 0.202 | 0.207 | 0.202 | 0.247 | 0.203 | 0.195 | 0.193 | 0.214 |
| R2 Adj. | 0.256 | 0.257 | 0.258 | 0.216 | 0.195 | 0.206 | 0.201 | 0.241 | 0.196 | 0.194 | 0.192 | 0.208 |

*Note:* Regression estimates are reported using different specifications across all three outcomes. The variable *Stand-Malfeasance* is a min-max standardized measure of the amount of malfeasance reported in the video, ranging from 0 to 1. The *Negative* columns report the *Magnitude* specification for respondents who received negative information (increasing malfeasance). In contrast, the *Positive* columns report the specifications for those who received positive information (i.e., declining malfeasance). Standard errors are clustered at the commune level. *$p < .05$, **$p < .01$, ***$p < .001$.

## A.4 Compliance

We assessed whether results hold under treatment compliance. Table A.8 shows that 1,831 subjects received the placebo video and 3,694 the treatment video via WhatsApp. While we cannot confirm whether recipients viewed the video due to WhatsApp's privacy rules, compliance is proxied by correct responses to a ChatBot attention check immediately afterward. Response rates were 59% in the placebo group and 56% in the treatment group.

The ChatBot survey included two attention checks: (1)*"What was the topic of the video?"*. Around 98% respondents answered correctly in the placebo group and 95% in the treatment; and

(2)*"What was the information reported in the video?"* answered correctly by 97% and 99%, respectively. Most respondents passed both checks. As noted in the main text, we define compliance based on correct answers to the second question: 59% in the placebo group and 56% in the treatment group.

A post-treatment survey was conducted 3–6 days after subjects received the videos. Among the 3,592 respondents, 2,249 were from the treatment group (61% response rate) and 1,343 from the placebo group (73%). Respondents were asked several attention checks. When asked whether they remembered receiving a video from Chile Transparente, 63% of placebo and 74% of treatment respondents said yes. Among those, 80% in the placebo group and 84% in the treatment group correctly recalled the video's content.

We defined "Post-Treatment Compliers" as the approximately 38% of subjects who answered the post-treatment survey and reported watching the treatment videos. We defined "Chatbot Compliers" as subjects who answered attention check questions in both the Chatbot and Post-treatment surveys. We used two definitions of compliance (Chatbot Compliers and Post-Treatment Compliers) to estimate the Complier Average Causal Effects (CACE).

Table A.9 reports pre-treatment differences between compliers and non-compliers. Compliers scored lower on the Resource Outcome measure and were more likely to donate compared to non-compliers. They are also more likely to be male and younger. In terms of political and socioeconomic characteristics, compliers tend to be more left-leaning and slightly wealthier than non-compliers. However, compliers and non-compliers are similar in their prior beliefs for the Qualitative and Distribution outcome and also similar in years of education and partisanship.

Table A.8: Treatment Compliance

| **Treatment status** | Placebo | % | Treatment | % | Total | % T. sample |
|---|---|---|---|---|---|---|
| Videos sent | 1832 | | 3696 | - | 5528 | 100 |
| **A:Chatbot survey** | | | | | | |
| Clicked video | No information | | | | | |
| Respondent Chatbot survey | 1077 | 100 | 2054 | 100 | 3131 | 57 |
| First compliance question - correct | 1055 | 98 | 1961 | 95 | 3016 | |
| First compliance question - incorrect | 22 | 2 | 112 | 5 | 134 | |
| Second compliance question - correct | 1042 | 97 | 2039 | 99 | 3130 | |
| Second compliance question - incorrect | 34 | 3 | 15 | 1 | 49 | |
| **B:Post-treatment survey** | | | | | | |
| Total post-treatment sample | 1832 | 100 | 3696 | 100 | 5528 | |
| Responded | 1343 | 73 | 2249 | 61 | 3592 | 65 |
| No response | 489 | 27 | 1447 | 39 | 1936 | |
| **C:Post-treatment - compliance** | | | | | | |
| Total post-treatment sample | 1343 | 100 | 2249 | 100 | 3592 | |
| Reported receiving the video | 846 | 63 | 1670 | 74 | 2516 | 46 |
| Attention check question - correct | 680 | 51 | 1401 | 62 | 2081 | |
| Attention check question - incorrect | 166 | 12 | 269 | 12 | 435 | |
| **D:Chatbot and Post-Treatment survey** | | | | | | |
| Total post-treatment sample | 1343 | 100 | 2249 | 100 | 3592 | |
| Responded both surveys | 987 | 73 | 2055 | 91 | 3042 | 55 |
| Attention check Chatbot correct | 967 | 72 | 1962 | 87 | 2929 | |
| Attention check Chatbot incorrect | 20 | 1 | 93 | 4 | 111 | |
| **F:Follow up survey - compliance** | | | | | | |
| Total follow-up sample | 1057 | 100 | 1198 | 100 | 2255 | |
| Reported receiving the video | 534 | 50 | 729 | 60 | 1263 | 23 |
| Attention check question - correct | 371 | 35 | 548 | 46 | 919 | |
| Attention check question - incorrect | 163 | 15 | 181 | 15 | 344 | |
| **G:Chatbot and Follow up survey** | | | | | | |
| Total follow-up sample | 1057 | 100 | 1198 | 100 | 2255 | |
| Reported receiving the video | 534 | 44 | 729 | 60 | 1263 | 22 |
| Attention check Chatbot correct | 411 | 39 | 559 | 45 | 970 | |
| Attention check Chatbot incorrect | 3 | 0.3 | 25 | 0.02 | 28 | |

*Note:* This table reports compliance and attention check results across three survey waves. Panel A displays the correct responses to attention checks in the Chatbot survey, which were asked immediately after the video. Panel B reports the completion of the post-treatment survey one week later. Panel C shows respondents who completed the post-treatment survey, watched the video, and answered the attention check. Panel D includes those who completed both surveys and passed both attention checks. Panel F summarizes follow-up survey compliance, and Panel G presents respondents who passed both Chatbot and post-treatment checks.

Table A.8 reports ”Post-Treatment Compliers” in Panel C, showing counts by treatment status, the percentage who reported receiving the videos, and those who answered the attention check correctly as a share of the post-treatment sample. Panel D presents analogous data for ”Chatbot Compliers” in both groups. Panels F and G present the same metrics for the follow-up survey. Panel F displays the proportion of post-treatment compliers, excluding follow-up respondents, while Panel G summarizes chatbot compliance among follow-up survey participants by treatment condition.

Table A.9: Pre-treatment Measures - Compliers Versus Non-Compliers

| | Baseline | Post-Treat | | | Follow-up | | |
|---|---|---|---|---|---|---|---|
| | Whole sample | Compliers | Non-compliers | Difference | Compliers | Non-compliers | Difference |
| **Qualitative Measures** | | | | | | | |
| Subjective malfeasance scale 1-10 | 4.49 | 4.44 | 4.46 | -0.02 | 4.45 | 4.38 | 0.08 |
| | (2.44) | (2.42) | (2.38) | | (2.39) | (2.34) | |
| Certainty Subjective malfeasance scale 1-10 | 5.40 | 5.43 | 5.31 | 0.13 | 5.25 | 5.49 | -0.24 |
| | (2.71) | (2.67) | (2.75) | | (2.66) | (2.68) | |
| **Quantitative Measures - Incentivized** | | | | | | | |
| Resources | 3.84 | 3.75 | 3.97 | -0.22 | 3.75 | 3.72 | 0.03 |
| | (2.38) | (2.33) | (2.48) | | (2.35) | (2.18) | |
| Distribution | 1.90 | 1.88 | 1.91 | -0.03 | 1.88 | 1.89 | -0.01 |
| | (0.78) | (0.79) | (0.78) | | (0.76) | (0.73) | |
| **Behavioural measures** | | | | | | | |
| 500 CLP Municipal Donation | 32% | 34% | 28% | 6% | 33% | 35% | -3% |
| | (0.63%) | (0.93%) | (1.38%) | | (1.33%) | (1.43%) | |
| **Covariates** | | | | | | | |
| Female | 64.73% | 63.77% | 65.9%2 | -2.15% | 67.34% | 66.41% | 0.93% |
| | (0.01%) | (0.96%) | (1.45%) | | (1.33%) | (1.51%) | |
| Education (Yrs) | 15.07 | 15.21 | 14.86 | 0.35 | 15.36 | 15.27 | 0.08 |
| | (0.04) | (0.06) | (0.09) | | (0.07) | (0.09) | |
| Age | 35.88 | 35.05 | 36.73 | -1.68 | 32.80 | 33.00 | -0.20 |
| | (0.18) | (0.26) | (0.43) | | (0.34) | (0.45) | |
| Partisanship | 4.47 | 4.34 | 4.61 | -0.27 | 4.21 | 4.39 | -0.18 |
| | (0.03) | (0.04) | (0.06) | | (0.06) | (0.06) | |
| Co-partisanship | 27.13 | 25.32 | 29.32 | -3.99 | 25.38 | 28.54 | -3.15 |
| | (0.60%) | (0.88%) | (1.38%) | | (1.25%) | (1.41%) | |
| Income (USD) | 658 | 646 | 606 | 40 | 521 | 652 | -131 |
| | (16) | (25) | (23) | | (19) | (32) | |
| Sample | 5528 | 2081 | 1511 | | 987 | 1268 | |

*Note:* This table reports pre-treatment values for all outcomes and relevant covariates broken down by compliance status. Along with mean values, we report standard deviations in parentheses. We also calculate the difference in means for each variable. Compliance status was determined by whether respondents reported receiving the video treatment in the post-treatment and follow surveys.

## A.5 Heterogeneity

**Education** Our *Information Treatment* model of the intent to treat subjects with the malfeasance information video results in a very robust and significant estimated treatment effect. This model implies a minimal engagement with the quantitative content of the message – beliefs update when subjects are simply reminded that there is malfeasance in local government. There is also support for our richer model in which belief updating responds to the quantitative content of the malfeasance information messages. This model presumes that subjects engage with the quantitative content of the video treatment. Hence, the likelihood that subjects will engage in the more demanding quantitative features of this information treatment will be conditioned on their cognitive ability. We use education as a proxy for cognitive ability. We have two expectations: We expect the treatment effect for the baseline model, with a dichotomous treatment variable, to be similar across education segments in the subject population. On the other hand, we conjecture that the higher educated are more responsive to the quantitative details of the malfeasance information treatment.

We exploit machine learning techniques in order to evaluate 1) whether education is a source of heterogeneous treatment effects and 2) how education-related heterogeneity varies by treatment. Our non-parametric modeling strategy uses Bayesian Additive Regression Trees (BART).

We recover ATE estimates by first using the BART models to generate predicted outcomes for the observed data and then for a set of counterfactual observations. To train the BART model, we provide the observed outcome of interest and a training data matrix consisting of the information treatment variable and the covariates of interest. We then simulate a second set of outcomes based on a separate test data matrix that contains "synthetic" observations that are identical to the training data except that the treatment assignment is inverted ('treated' becomes 'placebo' and vice versa). The test data are only used post-estimation to predict counterfactual outcomes based on the results of the BART model, which were trained on the observed training data. Estimating outcomes for *both* the observed and synthetic observations ensures that for each observation there is a corresponding counterfactual "placebo" case, where the only difference between the two is the value for treatment assignment. The individual-level effect of being treated is simply the difference between these two predictions.[3]. Conveniently, with individual-level predictions, we can assess the extent to which education, taking into consideration other covariates, is a principal source of heterogeneity in the malfeasance video treatment effect.[4]

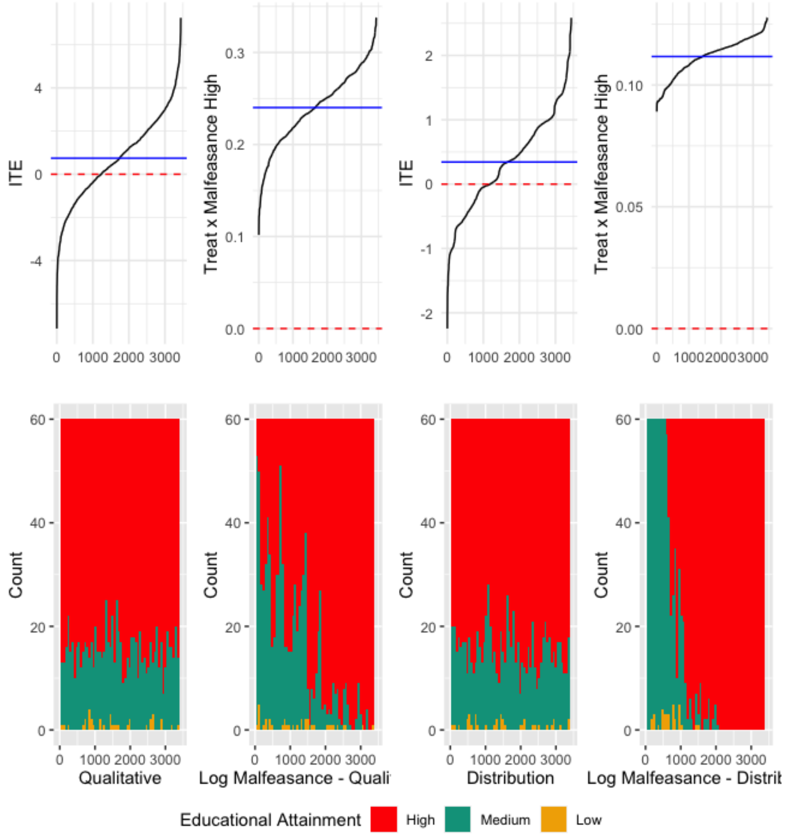
In the case of the Qualitative and Distribution outcomes, the results for BART analyses of the Information Treatment and Magnitudes treatment effect models confirm our conjectures. Figure A.4 plots the distribution of the estimated individual-level treatment effects by magnitude along with the histogram of the corresponding education covariate values beneath. For these two outcome variables, the Information Treatment model has a significant malfeasance treatment effect.

[3] Our BART model of heterogeneous effects is a simple specification generated using the BART R package with inputs described above. All other options within BART are left at their default value.

[4] Although there are clear data constraints here – we do not claim to have measured and estimated the effect of all possible covariates. Instead, our claim pertains to the various 'pre-treatment" variables collected as part of the surveys.

The plot in the top part of panel (a) of Figure A.4 clearly demonstrates that the size of this effect is heterogeneous in the sample population: the overall estimated ATE for the Qualitative measure (that varies between 1 and 10) is approximately 0.75 (identical to the estimated coefficient in Table 2) and it ranges between -7.2 to 7.2. The plot in the lower part of panel (a) of Figure A.4 indicates the density of the three education covariate values over the range of estimated CATEs. The distribution of CATEs generated by BART suggests that education is not correlated with the magnitude of the treatment effects. The effect of being exposed to an information treatment reporting audited malfeasance is very similar across education groups in the sampled population. Panel (b) of Figure A.4 reports similar BART analyses of the Information treatment effect for the Distribution outcome variable. Again, we find that exposure to our audit information treatment generates belief updating effects with similar density distributions across our three education groups. These results confirm our expectation that simple belief updating in response to any information about malfeasance would not be conditioned on educational levels.

Figure A.4: Heterogeneous Conditional Average Treatment Effects: Qualitative and Distribution Corruption Beliefs



*Note:* BART estimated heterogeneous effects by education; predicted CATEs are in the top panel of each column; and corresponding histograms of education distributions are in the bottom panel.

Our BART analysis does detect evidence, consistent with our conjecture, that the effect of malfeasance magnitudes would be conditional on education levels.[5] Results of the BART analyses of the malfeasance magnitude effect for both the Qualitative and Distribution outcome variables are reported in Panels (c) and (d) of Figure A.4. In both cases, the medium-educated subjects have CATEs that are skewed in the direction of the lower part of the continuum, while the higher-educated subjects are skewed in the higher part of the CATE continuum. The better educated are clearly much more responsive to quantitative information regarding malfeasance magnitudes.

When modeling the treatment as a simple binary variable, the effect is highly significant and

[5] To enable more straightforward estimation of the difference in predicted outcomes, we dichotomize the log of malfeasance variable into high (1) and low (0) respectively – we dichotomize at the median value of 8.0.

similar across education categories. This was our expectation given that the treatment effect does not imply any significant cognitive engagement with the quantitative content of the video. Belief updating is also correlated with malfeasance magnitudes that are described in the information treatments. This correlation we believe will be conditional on education because, we conjecture, the quantitative details of the information treatment require more cognitive engagement on the part of subjects. We find that the malfeasance magnitude treatment effect is conditioned on education levels. However, we observe no evidence of education-related heterogeneous treatment effects in the case of the Resources outcome variable.

## A.6 Corruption Audits in Chile

**Information about the audit process and malfeasance found in audits** The source of the information treatment in this experiment is the CGO annual audit program. The CGO typically focuses on government activities such as procurement, hiring processes, or municipal finances. Audited municipalities are selected based on a scoring system incorporating factors such as budget size, transfers to the private sector, and results from previous audits. This scoring system is neither publicly available nor accessible to municipalities. The factors and weights of each element change from one audit program to another.

The CGO audits reveal an average of 31 irregularities per municipality, amounting to USD 1,119 to 6,695,744, or roughly 4% of the average municipal budget in Chile. Irregularities are classified by severity, with about 58% categorized accordingly. Highly complex irregularities involve significant restitution requests or potential civil or criminal liabilities and affect critical processes, often triggering disciplinary actions by audit or government agencies. Complex irregularities indicate severe control weaknesses that may lead to investigations by other institutions. Moderately complex irregularities indicate procedural failures that do not impact critical processes and may not necessitate further action. Mild irregularities are minor and have no impact on critical processes or follow-up. We argue that highly complex irregularities correspond to outright corruption, while the others represent mismanagement, given the absence of criminal prosecution or restitution of funds.

The CGO (and other independent governmental institutions) can prosecute civil servants and mayors for the irregularities found in these audits. Civil servants or mayors found guilty of violating statutory rules or engaging in criminal activities are subject to disciplinary action, which may include reparation orders, unpaid leave, and more severe penalties, such as dismissals, demotions, fines, and non-custodial penalties. On average, 5 out 31 of the irregularities found in each audit involve some form of disciplinary action.

As noted earlier, our study takes place in a context where corruption is highly salient. Overall perceptions of corruption are high, with 88% of respondents in the *Consejo para la Transparencia*'s annual representative survey believing that more than half of Chilean politicians are involved in corrupt practices (para la Transparencia, 2020). However, when asked which government institutions exhibit high corruption levels, municipal governments rank among the lowest; fewer than 4% of respondents perceive corruption as widespread at the local level (para la Transparencia, 2019). In terms of direct experience, less than 3% of respondents reported being solicited for bribes by public officials (para la Transparencia, 2018, 2019, 2020).

## A.7 Additional Analyses, Subgroup Analyses and Persistence Ratios

Table A.10: Regression Results of Malfeasance Beliefs - Information Gap

| | Resources | Distribution |
|---|---|---|
| Intercept | 2.171*** | 0.990*** |
| | (0.654) | (0.156) |
| Prior | 0.502*** | 0.481*** |
| | (0.025) | (0.034) |
| Treat | 0.735*** | 0.327*** |
| | (0.096) | (0.030) |
| Corrup diff | 0.059* | 0.030 |
| | (0.030) | (0.034) |
| Treat:Corrup diff | 0.000 | 0.051* |
| | (0.028) | (0.024) |
| Num.Obs. | 3439 | 3439 |
| R2 | 0.223 | 0.207 |
| R2 Adj. | 0.216 | 0.200 |

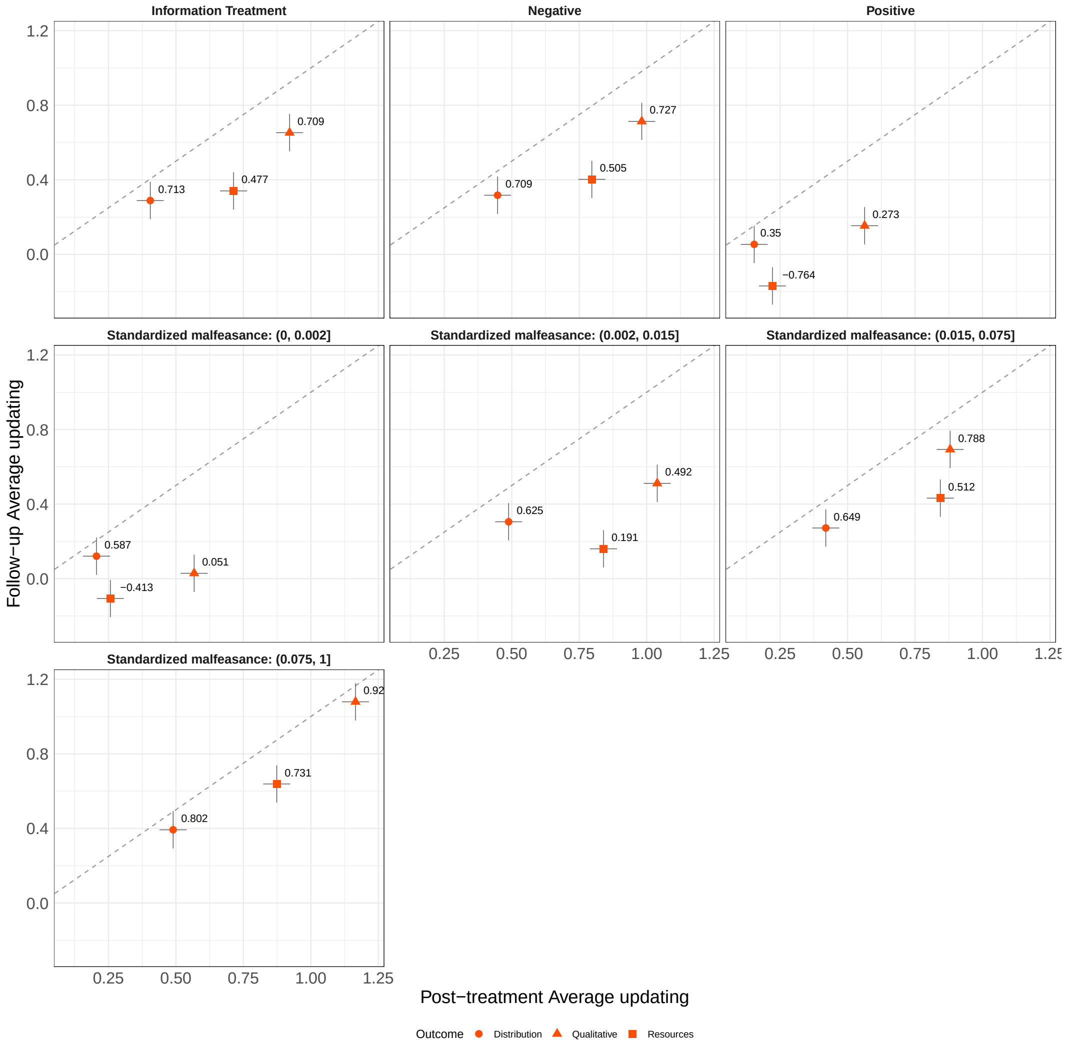
*Note:* This table reports regression estimates for the *Information Gap* model. Standard errors are clustered at the commune level. The Corrup diff variable captures the difference between respondents beliefs pre-treatment versus the ground true $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table A.11: Regression Results of Malfeasance Beliefs - Priors Group

| | Qualitative | | | Resources | | | Distribution | | |
|---|---|---|---|---|---|---|---|---|---|
| | Pooled | High | Low | Pooled | High | Low | Pooled | High | Low |
| **Panel A: Information** | | | | | | | | | |
| Intercept | 2.734*** | 2.330*** | 2.174*** | 2.302*** | 2.513** | 2.061* | 1.122*** | 0.267 | 1.059*** |
| | (0.401) | (0.480) | (0.611) | (0.662) | (0.825) | (0.865) | (0.153) | (0.567) | (0.145) |
| Prior | 0.467*** | 0.561*** | 0.516*** | 0.446*** | 0.355*** | 0.668*** | 0.421*** | 0.512*** | 0.499*** |
| | (0.020) | (0.043) | (0.030) | (0.018) | (0.035) | (0.054) | (0.024) | (0.058) | (0.034) |
| Treat | 0.749*** | 0.641*** | 0.824*** | 0.740*** | 0.912*** | 0.643*** | 0.347*** | 0.340*** | 0.351*** |
| | (0.079) | (0.129) | (0.096) | (0.088) | (0.151) | (0.098) | (0.028) | (0.060) | (0.032) |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Num.Obs. | 3439 | 1336 | 2103 | 3439 | 1384 | 2055 | 3439 | 770 | 2669 |
| R2 | 0.263 | 0.206 | 0.209 | 0.220 | 0.136 | 0.139 | 0.203 | 0.235 | 0.176 |
| R2 Adj. | 0.257 | 0.191 | 0.199 | 0.214 | 0.120 | 0.128 | 0.196 | 0.208 | 0.167 |
| **Panel B: Benchmarking** | | | | | | | | | |
| Intercept | 2.471*** | 1.762** | 2.107** | 2.198*** | 2.121** | 2.120* | 1.101*** | 0.161 | 1.109*** |
| | (0.446) | (0.575) | (0.647) | (0.642) | (0.821) | (0.867) | (0.166) | (0.599) | (0.163) |
| Prior | 0.466*** | 0.562*** | 0.515*** | 0.445*** | 0.358*** | 0.665*** | 0.420*** | 0.502*** | 0.498*** |
| | (0.020) | (0.043) | (0.029) | (0.019) | (0.035) | (0.063) | (0.026) | (0.057) | (0.039) |
| Treat | 0.575** | 0.665 | 0.539 | 0.321 | 1.014*** | -0.116 | 0.110 | 0.336 | 0.024 |
| | (0.208) | (0.369) | (0.293) | (0.167) | (0.287) | (0.260) | (0.067) | (0.193) | (0.076) |
| Negative | 0.311 | 0.618 | 0.109 | 0.132 | 0.443 | -0.129 | 0.031 | 0.249 | -0.067 |
| | (0.212) | (0.354) | (0.236) | (0.174) | (0.320) | (0.184) | (0.084) | (0.159) | (0.087) |
| Treat:Negative | 0.202 | -0.027 | 0.334 | 0.489* | -0.114 | 0.887** | 0.278*** | 0.003 | 0.380*** |
| | (0.227) | (0.401) | (0.310) | (0.190) | (0.327) | (0.279) | (0.073) | (0.205) | (0.082) |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Num.Obs. | 3439 | 1336 | 2103 | 3439 | 1384 | 2055 | 3439 | 770 | 2669 |
| R2 | 0.267 | 0.214 | 0.212 | 0.225 | 0.138 | 0.148 | 0.213 | 0.247 | 0.186 |
| R2 Adj. | 0.261 | 0.198 | 0.201 | 0.218 | 0.121 | 0.136 | 0.206 | 0.218 | 0.177 |
| **Panel C: Magnitudes** | | | | | | | | | |
| Intercept | 2.717*** | 3.098*** | 2.498*** | 2.274*** | 2.731*** | 1.737* | 1.125*** | 0.980*** | 0.969*** |
| | (0.402) | (0.506) | (0.569) | (0.648) | (0.788) | (0.868) | (0.152) | (0.227) | (0.194) |
| Prior | 0.464*** | 0.436*** | 0.426*** | 0.444*** | 0.342*** | 0.671*** | 0.415*** | 0.414*** | 0.578*** |
| | (0.020) | (0.033) | (0.024) | (0.018) | (0.035) | (0.053) | (0.024) | (0.041) | (0.035) |
| Treat | 0.736*** | 0.812*** | 0.698*** | 0.739*** | 0.899*** | 0.676*** | 0.327*** | 0.335*** | 0.337*** |
| | (0.091) | (0.128) | (0.127) | (0.094) | (0.165) | (0.123) | (0.035) | (0.053) | (0.040) |
| Stand - Malfeasance | 0.490 | 0.404 | 0.666 | 0.590 | 0.636 | 0.628 | 0.217 | 0.181 | 0.251 |
| | (0.489) | (0.664) | (0.894) | (0.432) | (0.617) | (0.842) | (0.233) | (0.291) | (0.264) |
| Treat:Stand - Malfeasance | 0.179 | -0.649 | 0.636 | 0.012 | -0.103 | -0.227 | 0.302 | 0.364 | 0.162 |
| | (0.882) | (0.502) | (1.263) | (0.597) | (0.985) | (1.414) | (0.353) | (0.530) | (0.354) |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Num.Obs. | 3439 | 1366 | 2073 | 3439 | 1366 | 2073 | 3439 | 1595 | 1844 |
| R2 | 0.264 | 0.223 | 0.226 | 0.221 | 0.129 | 0.138 | 0.207 | 0.180 | 0.222 |
| R2 Adj. | 0.258 | 0.207 | 0.215 | 0.214 | 0.111 | 0.126 | 0.200 | 0.165 | 0.210 |
| **Panel D: Information Gap** | | | | | | | | | |
| Intercept | | | | 2.171*** | 2.366** | 1.952* | 0.990*** | 0.121 | 0.990*** |
| | | | | (0.654) | (0.835) | (0.862) | (0.156) | (0.558) | (0.159) |
| Prior | | | | 0.502*** | 0.418*** | 0.701*** | 0.481*** | 0.594*** | 0.543*** |
| | | | | (0.025) | (0.042) | (0.055) | (0.034) | (0.066) | (0.046) |
| Treat | | | | 0.735*** | 0.992*** | 0.639*** | 0.327*** | 0.349*** | 0.288*** |
| | | | | (0.096) | (0.248) | (0.099) | (0.030) | (0.069) | (0.041) |
| Corrup diff | | | | 0.059* | 0.054 | 0.024 | 0.030 | 0.085 | -0.009 |
| | | | | (0.030) | (0.041) | (0.036) | (0.034) | (0.048) | (0.036) |
| Treat:Corrup diff | | | | 0.0003 | 0.022 | 0.033 | 0.051* | 0.008 | 0.089* |
| | | | | (0.028) | (0.052) | (0.044) | (0.024) | (0.055) | (0.035) |
| Covariates | | | | Yes | Yes | Yes | Yes | Yes | Yes |
| Num.Obs. | | | | 3439 | 1384 | 2055 | 3439 | 770 | 2669 |
| R2 | | | | 0.223 | 0.140 | 0.142 | 0.207 | 0.244 | 0.180 |
| R2 Adj. | | | | 0.216 | 0.122 | 0.129 | 0.200 | 0.216 | 0.171 |

*Note:* This table reports regression estimates by prior beliefs about malfeasance for three outcomes and belief groups ("Pooled", "High", "Low"). We use same models reported in Tables 2 and A.10. Standard errors clustered at the commune level. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Figure A.5: Persistence ratios

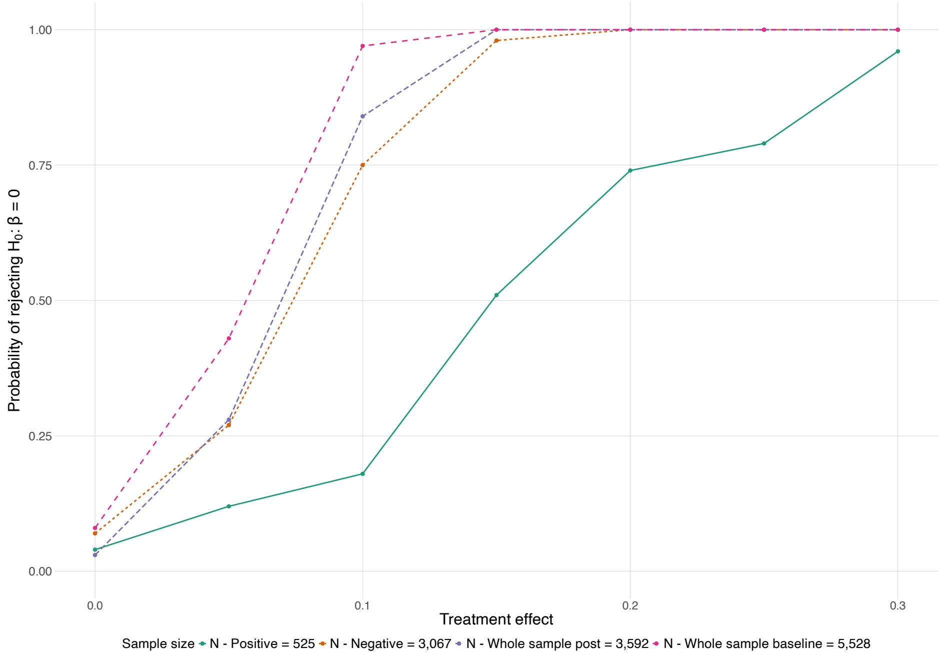


*Note:* This figure reports persistence ratios of average updating estimates for all three outcomes. Persistence ratios are calculated by dividing one-month estimates by estimates obtained 3-6 days after the treatment was delivered. A ratio of 1 would be equivalent to no treatment decay a month after the intervention. 3-6 days treatment effects (Post-treatment) are reported in the x-axis, while one-month estimates are reported in the y-axis. Persistence ratios are reported next to each crosshair point. Negative persistence ratios correspond to cases where there is a sign reversal between the post-treatment and follow-up estimates. We broke down persistence ratios by treatment status: ”Information Treatment,” ”Negative,” and ”Positive,” and different levels of standardized measures of malfeasance. The ”Negative” treatment condition comprises all respondents who received unfavorable information, whereas ”Positive” includes subjects who received a favorable information frame.

## A.8 Power calculations

We conducted power calculations using simulations to test the probability of rejecting the null hypothesis of no effect for different treatment effects. Based on the results of this simulation, we are well-powered for both the whole sample and the ”negative” frame sample. However, we are underpowered for the ”positive” frame.

Figure A.6: Power Calculations



*Note:* This figure displays the probability of rejecting the null hypothesis of no treatment effect across a range of effect sizes. Power calculations are based on the sample sizes used in the study.

## A.9 Ethics Approvals

This research project received ethics approval from the relevant institutions; details have been withheld to preserve anonymity during the review process. We adhered to the [Country] Data Protection Acts of 1998 and 2018, as well as Chile’s Data Protection Act (Law No. 19.628). All participants provided informed consent, and we ensured that their data was handled securely and in compliance with relevant data protection regulations. Participants were compensated approximately $3.70 per survey, and no deception was involved in the study. The University of —— served as the primary data controller

# References

Adida, C., Gottlieb, J., Kramon, E., & McClendon, G. (2020). When does information influence voters? the joint importance of salience and coordination. *Comparative Political Studies*, *53*(6), 851–891.

Anderson, K., Zamarro, G., Steele, J., & Miller, T. (2021). Comparing performance of methods to deal with differential attrition in randomized experimental evaluations. *Evaluation Review*, *45*(1-2), 70–104.

Anduiza, E., Gallego, A., & Muñoz, J. (2013). Turning a blind eye: Experimental evidence of partisan bias in attitudes toward corruption. *Comparative Political Studies*, *46*(12), 1664–1692.

Arel-Bundock, V., Blais, A., & Dassonneville, R. (2021). Do voters benchmark economic performance? *British Journal of Political Science*, *51*(1), 437–449. https://doi.org/10.1017/S0007123418000236

Arias, E., Balán, P., Larreguy, H., Marshall, J., & Querubín, P. (2019). Information provision, voter coordination, and electoral accountability: Evidence from mexican social networks. *American Political Science Review*, *113*(2), 475–498. https://doi.org/10.1017/S0003055419000091

Arias, E., Larreguy, H., Marshall, J., & Querubín, P. (2022a). Priors Rule: When Do Malfeasance Revelations Help Or Hurt Incumbent Parties? *Journal of the European Economic Association*, *20*(4), 1433–1477.

Arias, E., Larreguy, H., Marshall, J., & Querubín, P. (2022b). Priors rule: When do malfeasance revelations help or hurt incumbent parties? *Journal of the European Economic Association*, *20*(4), 1433–1477. https://doi.org/10.1093/jeea/jvac015

Arias, E., Larreguy, H. A., Marshall, J., & Querubín, P. (2025). Does the content and mode of delivery of information matter for electoral accountability? evidence from a field experiment in mexico. *Latin American Economic Review*, *34*, 1–41. https://doi.org/10.60758/laer.v34i.335

Avenburg, A. (2019). Public costs versus private gain: Assessing the effect of different types of information about corruption incidents on electoral accountability. *Journal of Politics in Latin America*, *11*(1), 71–108. https://doi.org/10.1177/1866802X19840457

Aytaç, S. E. (2018). Relative economic performance and the incumbent vote: A reference point theory. *The Journal of Politics*, *80*(1), 16–29. https://doi.org/10.1086/693908

Bandiera, O., Prat, A., & Valletti, T. (2009). Active and passive waste in government spending: Evidence from a policy experiment. *American Economic Review*, *99*(4), 1278–1308. https://doi.org/10.1257/aer.99.4.1278

Banerjee, A., Green, D. P., McManus, J., & Pande, R. (2014). Are poor voters indifferent to whether elected leaders are criminal or corrupt? a vignette experiment in rural india. *Political Communication*, *31*(3), 391–407. https://doi.org/10.1080/10584609.2014.914615

Barro, R. (1973). The control of politicians: An economic model. *Public Choice*, *14*(1), 19–42. https://doi.org/10.1007/BF01718440

Bauhr, M., & Charron, N. (2018). Insider or outsider? grand corruption and electoral accountability. *Comparative Political Studies*, *51*(4), 415–446. https://doi.org/10.1177/0010414017710258

Becher, M., Brouard, S., & Stegmueller, D. (2023). Endogenous benchmarking and government accountability: Experimental evidence from the covid-19 pandemic. *British Journal of Political Science*, 1–18. https://doi.org/10.1017/S0007123423000170

Beekman, G., Bulte, E., & Nillesen, E. (2014). Corruption, investments and contributions to public goods: Experimental evidence from rural liberia. *Journal of Public Economics*, *115*, 37–47.

Berliner, D., & Wehner, J. (2022). Audits for accountability: Evidence from municipal by-elections in south africa. *The Journal of Politics*, *84*(3), 1581–1594. https://doi.org/10.1086/716951

Bhandari, A., Larreguy, H., & Marshall, J. (2019). *Able and mostly willing: An empirical anatomy of information's effect on voter-driven accountability in 7yhb77uhn mvk,87-* [Working paper].

Blinder, S., & Schaffner, B. F. (2019). Going with the flows: Information that changes americans' immigration preferences. *International Journal of Public Opinion Research*, *32*(1), 153–164. https://doi.org/10.1093/ijpor/edz007

Boas, T. C., Hidalgo, F. D., & Melo, M. A. (2019). Norms versus action: Why voters fail to sanction malfeasance in brazil. *American Journal of Political Science*, *63*(2), 385–400. https://doi.org/10.1111/ajps.12413

Bobonis, G. J., Cámara Fuertes, L. R., & Schwabe, R. (2016). Monitoring corruptible politicians. *American Economic Review*, *106*(8), 2371–2405. https://doi.org/10.1257/aer.20130874

Bonica, A. (2018). Are donation-based measures of ideology valid predictors of individual-level policy preferences? *The Journal of Politics*, *81*. https://doi.org/10.1086/700722

Botero, S., Cornejo, R. C., Gamboa, L., Pavao, N., & Nickerson, D. W. (2015). Says who? an experiment on allegations of corruption and credibility of sources. *Political Research Quarterly*, *68*(3), 493–504. https://doi.org/10.1177/1065912915591607

Breitenstein, S. (2019). Choosing the crook: A conjoint experiment on voting for corrupt politicians. *Research & Politics*, *6*(1), 2053168019832230. https://doi.org/10.1177/2053168019832230

Brollo, F., Nannicini, T., Perotti, R., & Tabellini, G. (2013). The political resource curse. *American Economic Review*, *103*(5), 1759–96. https://doi.org/10.1257/aer.103.5.1759

Buckley, N., Reuter, O. J., Rochlitz, M., & Aisin, A. (2022). Staying out of trouble: Criminal cases against russian mayors. *Comparative Political Studies*, *55*(9), 1539–1568. https://doi.org/10.1177/00104140211047399

Buntaine, M. T., Jablonski, R., Nielson, D. L., & Pickering, P. M. (2018). Sms texts on corruption help ugandan voters hold elected councillors accountable at the polls. *Proceedings of the National Academy of Sciences*, *115*(26), 6668–6673. https://doi.org/10.1073/pnas.1722306115

Campos-Vazquez, R. M., & Mejia, L. A. (2016). Does corruption affect cooperation? a laboratory experiment. *Latin American Economic Review*, *25*(1), 5. https://doi.org/10.1007/s40503-016-0035-0

Carey, J. M., Chun, E., Cook, A., Fogarty, B. J., Jacoby, L., Nyhan, B., Reifler, J., & Sweeney, L. (2025). The narrow reach of targeted corrections: No impact on broader beliefs about election integrity. *Political Behavior*, *47*(2), 737–750. https://doi.org/10.1007/s11109-024-09968-0

Cavallo, A., Cruces, G., & Perez-Truglia, R. (2017). Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics*, *9*(3), 1–35. https://doi.org/10.1257/mac.20150147

Charbonneau, É., & Van Ryzin, G. G. (2015). Benchmarks and citizen judgments of local government performance: Findings from a survey experiment. *Public Management Review*, *17*(2), 288–304. https://doi.org/10.1080/14719037.2013.798027

Cheeseman, N., & Peiffer, C. (2020). *The unintended consequences of anti-corruption messaging in nigeria: Why pessimists are always disappointed* (tech. rep.). Anti-Corruption Evidence SOAS Consortium.

Cheeseman, N., & Peiffer, C. (2021). The curse of good intentions: Why anticorruption messaging can encourage bribery. *American Political Science Review*, 1–15. https://doi.org/10.1017/S0003055421001398

Chesseman, N., & Peiffer, C. (2024). Anti- corruption awareness raising campaigns: Why do they fail, and how can "backfire" effects be avoided? In *Handbook of anti-corruption research and practice*. Routledge.

Chong, A., De La O, A. L., Karlan, D., & Wantchekon, L. (2015). Does corruption information inspire the fight or quash the hope? a field experiment in mexico on voter turnout, choice, and party identification. *The Journal of Politics*, *77*(1), 55–71.

Coibion, O., Georgarakos, D., Gorodnichenko, Y., & van Rooij, M. (2023). How does consumption respond to news about inflation? field evidence from a randomized control trial. *Amer-*

*ican Economic Journal: Macroeconomics*, *15*(3), 109–52. https://doi.org/10.1257/mac.20200445

Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, *105*(8), 2644–78. https://doi.org/10.1257/aer.20110306

Coppock, A. (2023). *Persuasion in parallel*. University of Chicago Press.

Corbacho, A., Gingerich, D. W., Oliveros, V., & Ruiz-Vega, M. (2016). Corruption as a self-fulfilling prophecy: Evidence from a survey experiment in costa rica. *American Journal of Political Science*, *60*(4), 1077–1092. https://doi.org/10.1111/ajps.12244

Cornejo, R. C. (2022). Same scandal, different interpretations: Politics of corruption, anger, and partisan bias in mexico. *Journal of Elections, Public Opinion and Parties*, *0*(0), 1–22.

Coutts, A. (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, *22*(2), 369–395. https://doi.org/10.1007/s10683-018-9572-5

Cubel, M., Papadopoulou, A., & Sánchez-Pagés, S. (2024). Identity and political corruption: A laboratory experiment. *Economic Theory*.

D'Acunto, F., & Weber, M. (2024). Why survey-based subjective expectations are meaningful and important. *Annual Review of Economics*, *16*(Volume 16, 2024), 329–357. https://doi.org/https://doi.org/10.1146/annurev-economics-091523-043659

de Figueiredo, M. F., Hidalgo, F. D., & Kasahara, Y. (2022). When do voters punish corrupt politicians? experimental evidence from a field and survey experiment. *British Journal of Political Science*, 1–12. https://doi.org/10.1017/S0007123421000727

de Sousa, L., Clemente, F., & Maciel, G. G. (2023). Mapping conceptualisations and evaluations of corruption through survey questions: Five decades of public opinion-centred research. *European Political Science*, *22*(3), 368–383. https://doi.org/10.1057/s41304-023-00422-z

DellaVigna, S., & Gentzkow, M. (2010). Persuasion: Empirical evidence. *Annual Review of Economics*, *2*(1), 643–669.

Denly, M. (2022). Measuring corruption using governmental audits: A new framework and dataset [Accessed = 2022-11-01]. https://mikedenly.com/research/audit-measurement

Duch, R. M., Loewen, P., Robinson, T. S., & Zakharov, A. (2025). Governing in the face of a global crisis: When do voters punish and reward incumbent governments? *Proceedings of the National Academy of Sciences*, *122*(4), e2405021122. https://doi.org/10.1073/pnas.2405021122

Dunning, T., Grossman, G., Humphreys, M., Hyde, S., McIntosh, C., & Nellis, G. (2019). *Information, accountability, and cumulative learning: Lessons from metaketa i*. Cambridge University Press.

Dunning, T., Grossman, G., Humphreys, M., Hyde, S. D., McIntosh, C., Nellis, G., Adida, C. L., Arias, E., Bicalho, C., Boas, T. C., Buntaine, M. T., Chauchard, S., Chowdhury, A., Gottlieb, J., Hidalgo, F. D., Holmlund, M., Jablonski, R., Kramon, E., Larreguy, H., . . . Sircar, N. (2019). Voter information campaigns and political accountability: Cumulative findings from a preregistered meta-analysis of coordinated trials. *Science Advances*, *5*(7).

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, *3*(2), 114–38. https://doi.org/10.1257/mic.3.2.114

Elia, E., & Schwindt-Bayer, L. A. (2022). Corruption perceptions, opposition parties, and reelecting incumbents in latin america. *Electoral Studies*, *80*, 102545.

Enríquez, J. R., Larreguy, H., Marshall, J., & Simpser, A. (2024). Mass political information on social media: Facebook ads, electorate saturation, and electoral accountability in mexico. *Journal of the European Economic Association*, *22*(4), 1678–1722. https://doi.org/10.1093/jeea/jvae011

Erlich, A., Gans-Morse, J., & Nichter, S. (2025). Selective bribery: When do citizens engage in corruption? *Comparative Political Studies*, *58*(5), 996–1036. https://doi.org/10.1177/00104140241259444

Esberg, J., & Mummolo, J. (2018). *Explaining misperceptions of crime* [Unpublished manuscript].

Fan, T. Q., Liang, Y., & Peng, C. a. (2024). *The inference-forecast gap in belief updating* [Revise and resubmit at *Econometrica*. Latest draft and experimental instructions available online]. https://www.dropbox.com/scl/fi/7gyxuu80maruqbdqjo3lv/ifgap.pdf

Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice*, *50*, 5–25.

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

Ferraz, C., & Finan, F. (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review*, *101*(4), 1274–1311. https://doi.org/10.1257/aer.101.4.1274

Fiorina, M. P. (1981). *Retrospective voting in american national elections*. Yale University Press.

Gagliarducci, S., & Manacorda, M. (2020). Politics in the family: Nepotism and the hiring decisions of italian firms. *American Economic Journal: Applied Economics*, *12*(2), 67–95. https://doi.org/10.1257/app.20170778

Gerber, A. S., & Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company, Inc.

Ghanem, D., Hirshleifer, S., & Ortiz-Beccera, K. (2023). Testing attrition bias in field experiments. *Journal of Human Resources*. https://doi.org/10.3368/jhr.0920-11190R2

Gill, J., & Walker, L. D. (2005). Elicited priors for bayesian model specifications in political science research. *Journal of Politics*, *67*(3), 841–872. https://doi.org/https://doi.org/10.1111/j.1468-2508.2005.00342.x

Gillen, B., Snowberg, E., & Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, *127*(4), 1826–1863.

Gonzalez-Ocantos, E., de Jonge, C. K., & Nickerson, D. W. (2014). The conditionality of vote buying norms: Experimental evidence from latin america. *American Journal of Political Science*, *58*(1), 197–211.

Greifer, N. (2022). *Cobalt: Covariate balance tables and plots. r package version 4.4.1* (tech. rep.).

Haaland, I., & Roth, C. (2021). Beliefs about racial discrimination and support for pro-black policies. *The Review of Economics and Statistics*, 1–38. https://doi.org/10.1162/rest_a_01036

Haaland, I., Roth, C., & Wohlfart, J. (2023). Designing information provision experiments. *Journal of Economic Literature*, *61*(1), 3–40. https://doi.org/10.1257/jel.20211658

Healy, A., & Lenz, G. S. (2014). Substituting the end for the whole: Why voters respond primarily to the election-year economy. *American Journal of Political Science*, *58*(1), 31–47. https://doi.org/https://doi.org/10.1111/ajps.12053

Hicken, A., Leider, S., Ravanilla, N., & Yang, D. (2015). Measuring vote-selling: Field evidence from the philippines. *American Economic Review*, *105*(5), 352–56. https://doi.org/10.1257/aer.p20151033

Hopkins, D. J., Sides, J., & Citrin, J. (2019). The muted consequences of correct information about immigration. *The Journal of Politics*, *81*(1), 315–320.

Jablonski, R. S., Buntaine, M. T., Nielson, D. L., & Pickering, P. M. (2021). Individualized text messages about public services fail to sway voters: Evidence from a field experiment on ugandan elections. *Journal of Experimental Political Science*, 1–13.

Jäger, S., Roth, C., Roussille, N., & Schoefer, B. (2024). Worker beliefs about outside options*. *The Quarterly Journal of Economics*, *139*(3), 1505–1556. https://doi.org/10.1093/qje/qjae001

Jahnke, B., & Weisser, R. A. (2019). How does petty corruption affect tax morale in sub-saharan africa? *European Journal of Political Economy*, *60*, 101751.

Kayser, M. A., & Peress, M. (2012). Benchmarking across borders: Electoral accountability and the necessity of comparison. *American Political Science Review*, *106*(3), 661–684. https://doi.org/10.1017/S0003055412000275

Kovach, M. (2021, January). *Conservative Updating* (Papers No. 2102.00152). arXiv.org. https://ideas.repec.org/p/arx/papers/2102.00152.html

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Kuziemko, I., Norton, M. I., Saez, E., & Stantcheva, S. (2015). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review*, *105*(4), 1478–1508. https://doi.org/10.1257/aer.20130360

Lagunes, P. (2021, September). *The Eye and the Whip: Corruption Control in the Americas*. Oxford University Press. https://doi.org/10.1093/oso/9780197577622.001.0001

LAPOP. (2021). Chile: Americasbarometer 2021 [Vanderbilt University, Nashville, TN].

Larreguy, H., Marshall, J., & Snyder, J., James M. (2020). Publicising Malfeasance: When the Local Media Structure Facilitates Electoral Accountability in Mexico. *The Economic Journal*, *130*(631), 2291–2327. https://doi.org/10.1093/ej/ueaa046

Larsen, M. V., & Olsen, A. L. (2020). Reducing bias in citizens' perception of crime rates: Evidence from a field experiment on burglary prevalence. *The Journal of Politics*, *82*(2), 747–752. https://doi.org/10.1086/706595

Letki, N., Górecki, M. A., & Gendźwiłł, A. (2023). 'they accept bribes; we accept bribery': Conditional effects of corrupt encounters on the evaluation of public institutions. *British Journal of Political Science*, *53*(2), 690–697. https://doi.org/10.1017/S0007123422000047

Liang, Y. (2025). Learning from unknown information sources. *Management Science*, *71*(5), 3873–3890. https://doi.org/10.1287/mnsc.2021.03551

Martinelli, C. (2022). Accountability and grand corruption. *American Economic Journal: Microeconomics*, *14*(4), 645–79. https://doi.org/10.1257/mic.20200186

Mian, A., Sufi, A., & Khoshkhou, N. (2021). Partisan Bias, Economic Expectations, and Household Spending. *The Review of Economics and Statistics*, 1–46. https://doi.org/10.1162/rest_a_01056

Mutz, D. C. (2021). *Winners and losers: The psychology of foreign trade* (Vol. 27). Princeton University Press. https://doi.org/10.2307/j.ctv1fj85hw

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

OECD. (2016). *Supreme audit institutions and good governance*.

para la Transparencia, C. N. (2018). Estudio nacional de transparencia.

para la Transparencia, C. N. (2019). Estudio nacional de transparencia.

para la Transparencia, C. N. (2020). Estudio nacional de transparencia.

Peiffer, C. (2020). Message received? experimental findings on how messages about corruption shape perceptions. *British Journal of Political Science*, *50*(3), 1207–1215. https://doi.org/10.1017/S0007123418000108

Peterson, E., & Iyengar, S. (2021). Partisan gaps in political information and information-seeking behavior: Motivated reasoning or cheerleading? *American Journal of Political Science*, *65*(1), 133–147.

Peyton, K. (2020). Does trust in government increase support for redistribution? evidence from randomized survey experiments. *American Political Science Review*.

Rathje, S., Roozenbeek, J., Bavel, J. J. V., & van der Linden, S. (2023). Accuracy and social motivations shape judgements of (mis)information. *Nature Human Behaviour*, *7*(6), 892–903. https://doi.org/10.1038/s41562-023-01540-w

Reinikka, R., & Svensson, J. (2006). Using micro-surveys to measure and explain corruption [Part Special Issue (pp. 324–404). Corruption and Development: Analysis and Measurement]. *World Development*, *34*(2), 359–370. https://doi.org/https://doi.org/10.1016/j.worlddev.2005.03.009

Ridgeway, G., McCaffrey, D., Morral, A., Cefalu, M., Burgette, L., Pane, J., & Griffin, B. A. (2021, October). *Toolkit for weighting and analysis of nonequivalent groups: A guide to the twang package* (R Package). RAND. https://cran.r-project.org/web/packages/twang/vignettes/twang.pdf

Rodríguez, I., Rodon, T., Unan, A., Herbig, L., Klüver, H., & Kuhn, T. (2025). Benchmarking pandemic response: How the uk's covid-19 vaccine rollout impacted diffuse and specific support for the eu. *British Journal of Political Science*, *55*, e35. https://doi.org/10.1017/S0007123424000802

Rodríguez Chatruc, M., Stein, E., & Vlaicu, R. (2021). How issue framing shapes trade attitudes: Evidence from a multi-country survey experiment. *Journal of International Economics*, *129*, 103428. https://doi.org/https://doi.org/10.1016/j.jinteco.2021.103428

Roth, C., Settele, S., & Wohlfart, J. (2022). Beliefs about public debt and the demand for government spending [Annals Issue: Subjective Expectations  Probabilities in Economics]. *Journal of Econometrics*, *231*(1), 165–187. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.09.011

Roth, C., & Wohlfart, J. (2020). How do expectations about the macroeconomy affect personal expectations and behavior? *Review of Economics and Statistics*, *102*(4), 731–748. https://doi.org/10.1162/rest_a_00867

Sanna, G. A., & Lagnado, D. (2025). Belief updating in the face of misinformation: The role of source reliability. *Cognition*, *258*, 106090. https://doi.org/https://doi.org/10.1016/j.cognition.2025.106090

Solaz, H., Vries, C. E. D., & de Geus, R. A. (2019). In-group loyalty and the punishment of corruption. *Comparative Political Studies*, *52*(6), 896–926.

Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology*, *66*(8), S84–S90.e1.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755–769. https://doi.org/https://doi.org/10.1111/j.1540-5907.2006.00214.x

Tappin, B. M., Pennycook, G., & Rand, D. G. (2020). Bayesian or biased? analytic thinking and political belief updating [Epub 2020 Jun 24]. *Cognition*, *204*, 104375. https://doi.org/10.1016/j.cognition.2020.104375

Thaler, M. (2024). The fake news effect: Experimentally identifying motivated reasoning using trust in news. *American Economic Journal: Microeconomics*, *16*(2), 1–38. https://doi.org/10.1257/mic.20220146

Transparency International. (2023). *Corruption perceptions index 2022* (Accessed: 14 May 2025). Transparency International. https://images.transparencycdn.org/images/Report_CPI2022_English.pdf

World Bank. (2021, July). Supreme audit institutions independence index: 2021 global synthesis report.

Zappalà, G. (2024). Adapting to climate change accounting for individual beliefs. *Journal of Development Economics*, *169*, 103289. https://doi.org/https://doi.org/10.1016/j.jdeveco.2024.103289

Zhou, F., & Oostendorp, R. (2014). Measuring true sales and underreporting with matched firm-level survey and tax office data. *The Review of Economics and Statistics*, *96*(3), 563–576. https://doi.org/10.1162/REST_a_00408

Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review*, *110*(2), 337–61. https://doi.org/10.1257/aer.20180728