



R I D G E

RESEARCH INSTITUTE FOR
DEVELOPMENT, GROWTH
AND ECONOMICS

SINCE 2014



Unraveling the Paradox of Anticorruption Messaging: Experimental Evidence from a Tax Administration Reform

Nicolas Ajzenman, Martín Ardanaz,
Guillermo Cruces, Germán Feierherd,
Ignacio Lunghi

RIDGE - LACEA Working Paper 012

political

**This paper was presented at the 2025 RIDGE May Forum
jointly organized with the LACEA Networks**

RIDGE-LACEA Working Papers are not formal publications. They are preliminary versions of research presented at the RIDGE Forum and posted on the RIDGE website with the author's permission. They remain the intellectual property of the authors and may be submitted for publication in academic journals.

Unraveling the Paradox of Anticorruption Messaging: Experimental Evidence from a Tax Administration Reform

Nicolas Ajzenman, *McGill University*

Martín Ardanaz, *Inter-American Development Bank*

Guillermo Cruces, *U. of Nottingham & CEDLAS-IIE-UNLP*

Germán Feierherd, *Universidad de San Andrés*

Ignacio Lunghi, *CEDLAS-IIE-UNLP*

Abstract

Corruption and its perception pose significant obstacles to development, as they undermine government capabilities, notably through reduced willingness to pay taxes among citizens. Research shows that prior beliefs systematically bias how individuals process information, which means efforts to combat corruption and improve public perceptions face a critical challenge: citizens with entrenched pessimistic beliefs about widespread corruption often resist change, reinforcing their preconceptions. We propose that providing an external benchmark of corruption to shift the reference point before delivering anticorruption messages can mitigate these backfiring effects. In a survey experiment exploiting an institutional reform to address corruption within Honduras's tax agency, we find that while corruption-focused messages often backfire, our sequential approach reduces perceived corruption and lowers tax evasion intentions. A government-led field experiment with over 30,000 taxpayers validates these results: targeted anticorruption messages increased tax compliance. Our findings highlight the importance of addressing pre-existing beliefs in designing effective anti-corruption campaigns.

Keywords: Corruption, Tax Administration, Tax Evasion, Survey Experiment.

1. Introduction

Corruption poses a significant obstacle to development, undermining economic growth (Mauro, 1995; Fisman and Svensson, 2007; Cieřlik and Goczek, 2018) and the efficient allocation of resources (Acemoglu and Verdier, 2000; Olken, 2006). Beyond its direct effects, the perception of corruption presents an equally critical challenge. When governments are perceived as corrupt, institutional trust is eroded (Daniele et al., 2023), corrupt social norms are reinforced (Ajzenman, 2021; Gulino and Masera, 2023) and citizens’ willingness to contribute to public goods is reduced, thereby weakening government capabilities (Levi, 1988; Besley and Mueller, 2021). This effect is particularly pronounced in the context of tax compliance, where perceptions of corruption or illegitimacy within tax authorities significantly reduce citizens’ adherence to their tax obligations (Levi et al., 2009; Weigel and Kabue Ngindu, 2023; Yamou et al., 2024). Thus, efforts targeted at improving the competence of tax administration officials are essential for strengthening fiscal capacity in low-capacity settings (Alm et al., 2016; Okunogbe and Tourek, 2024).

Reforms to combat corruption and improve public perceptions of key institutions face a critical challenge: citizens with entrenched beliefs about widespread corruption often resist change, reinforcing their preconceptions. A substantial body of literature demonstrates that prior beliefs systematically bias how individuals process information and form judgments (e.g., Kunda, 1990; Zaller, 1992; Taber and Lodge, 2006). Theoretical frameworks such as motivated reasoning (Taber and Lodge, 2006), Bayesian updating (Bullock, 2009; Christensen, 2023; Broockman and Kalla, 2023), the Receive-Accept-Sample (RAS) model (Zaller, 1992), and priming (Iyengar and Kinder, 2010) all emphasize the pivotal role of prior beliefs in mediating responses to new information.¹ Despite their differences, these frameworks converge on a key insight: priors are not merely passive filters but active determinants of how individuals interpret and respond to new information.

The “backfire” effect, where information about government action inadvertently reinforces negative perceptions—such as the belief that corruption is pervasive—has been

¹E.g., see Cotter et al. (2020b) and Guess and Coppock (2020).

documented in various studies (Peiffer, 2020; Cheeseman and Peiffer, 2022), although recent research offers a more optimistic view, suggesting that this effect may not always occur, especially when individuals perceive the messages as credible (Guess and Coppock, 2020; Erlich and Gans-Morse, 2025). Evidence from Latin America highlights mostly negative impacts of anticorruption messages, with only a few cases showing positive effects when framed effectively (Corbacho et al., 2016; Beesley and Hawkins, 2022; Agerberg, 2022). This phenomenon is particularly likely when individuals hold deeply entrenched pessimistic priors and are influenced by confirmation or disconfirmation biases. For example, a seemingly positive statement like, “The government implemented a policy to reduce corruption,” can be interpreted very differently depending on one’s pre-existing beliefs. In highly polarized or skeptical contexts, individuals with negative priors may view such efforts as superficial or disingenuous, reinforcing their belief that corruption remains entrenched and that government initiatives are ineffective.

A promising approach to enhance the effectiveness of information-based anti-corruption interventions is therefore sequential: first improve priors about the extent of corruption, and only then provide information about specific government anti-corruption actions. Citizens may be more receptive to evidence of successful anti-corruption efforts if their pessimistic priors of corruption are attenuated beforehand.

We test this hypothesis leveraging a major institutional reform within Honduras’s national tax administration (*Servicio de Administración de Rentas* or SAR), drawing on data from both a survey experiment and a field experiment with actual taxpayers. The reform involved strengthening the tax administration by improving human resource policies and renewing the majority of existing personnel to address corruption within the tax agency. We test three types of informational treatments. A “reform” message informing individuals about recent efforts by the government to address corruption within the tax administration. A “debiasing” message, using a purposefully neutral non-government source (Latinobarometro) to show that, in the context of Latin America, Honduras’ tax administration corruption was indeed among the lowest in the region, thereby shifting individuals’ reference points. Finally, a “sequential” message, in which we first show

the “debiasing” message and then the “reform” one. Evidence from Yair et al. (2020) demonstrates that anticorruption measures are most effective when they come from credible sources, as they make integrity considerations more salient. Building on this, we hypothesize that highlighting such measures in the context of tax administration could influence taxpayers’ willingness to comply by strengthening perceptions of institutional integrity. We examined the effect of each of these messages on individuals’ beliefs about corruption within the tax authority, their general attitudes towards corruption, and their propensity for tax evasion in a survey experiment. We then analyze a government-led field experiment using similar messages on reported taxable income across over 30,000 tax returns.

Our findings align with theoretical predictions. Participants in our sample overwhelmingly perceived corruption among public officials as widespread, with a mean and median score of 8 and 9, respectively, on a 10-point scale. Consistent with prior research, the “reform” treatment did not improve respondents’ corruption beliefs or attitudes. On the contrary, it marginally reinforced negative perceptions of corruption levels, although these effects were generally non-significant ($p\text{-value} = 0.119$). In contrast, the “debiasing” message led to significant improvements in corruption beliefs. Finally, the “sequential” treatment, which delivered the “reform” message immediately after the “debiasing” message, generated strong and significant improvements in corruption beliefs, surpassing the “reform” treatment effect (significant at 1%).

Improvements in perceptions of corruption of the tax authority are expected to increase tax compliance (Alm et al., 2016; Martin, 2023; Baum et al., 2024). To evaluate the effects of our treatments on taxpayers’ willingness to evade taxes, we employed a crosswise model designed to mitigate social desirability bias. This method pairs a sensitive question about tax evasion with a neutral statement about the respondent’s mother’s birth month. By leveraging the known distribution of birth months, the approach estimates the true prevalence of tax evasion intentions without requiring direct responses (Jann et al., 2012). The “reform” message had positive but statistically insignificant effects on the willingness to evade taxes, while the “debiasing” message produced negative

yet similarly insignificant effects. In contrast, the “sequential” treatment demonstrated a significant negative effect on tax evasion intentions. However, a self-reported measure of tax evasion willingness, more prone to social desirability bias, showed no variation across treatment groups, contrary to our expectations.

While the survey results generally supported our hypotheses, the ultimate goal of these interventions is to influence actual tax behavior. To assess real-world effects, we extended our analysis to a government-led field experiment based on the survey findings. The government randomly assigned taxpayers to receive messages similar to those in the survey experiment in the weeks prior to the income tax filing deadline. Due to concerns about the “reform” treatment potentially backfiring, the field experiment included only the “debiasing” and “sequential” treatments alongside a control group. The field experiment findings align with the survey results. The “sequential” treatment significantly increased declared taxable income, while the “debiasing” treatment showed a positive but modest and statistically insignificant effect. Notably, the “sequential” treatment increased declared income by 6.8% overall and by 11.5% among higher-risk taxpayers, highlighting its efficacy in improving tax compliance.

Our study contributes to several strands of literature. First, our paper is related to the literature that highlights the unintended negative consequences of anticorruption messages, particularly when they amplify negative perceptions of government corruption or inefficacy (Chong et al., 2015; Corbacho et al., 2016; Peiffer, 2020; Beesley and Hawkins, 2022; Cheeseman and Peiffer, 2022). We contribute to this literature by showing that well-designed information campaigns, which consider prior beliefs and address perceptions of corruption before highlighting government reforms, can modify beliefs about the extent of corruption. Thus, our findings reinforce the idea that backfiring or backlash effects need not be as pervasive as previously suggested, and in some cases, may not occur at all when messages are perceived as credible (Guess and Coppock, 2020; Erlich and Gans-Morse, 2025).

Our paper is also connected to research on belief formation anomalies (DellaVigna, 2009; Cotter et al., 2020a). Individuals often deviate from Bayesian learning due to fac-

tors such as confirmation bias (Rabin and Schrag, 1999; Taber and Lodge, 2006; Little et al., 2022), motivated reasoning (Taber et al., 2009; Bénabou and Tirole, 2016; Epley and Gilovich, 2016; Adida et al., 2017; Little et al., 2022), and mistrust in information sources (Flynn et al., 2017). The failure to update beliefs can significantly impact critical outcomes, including assessments of government performance (Enríquez et al., 2024), responses to partisan cues (Bullock, 2011; Nicholson, 2012; Boudreau and MacKenzie, 2014), and the success of fact-checking efforts (Nyhan and Reifler, 2010; Barrera et al., 2020). By linking these belief formation anomalies to concrete behavioral outcomes, our findings shed light on the mechanisms through which interventions targeting beliefs can enhance governance and encourage contributions to public goods.

Finally, our paper is related to a growing body of empirical literature that uses field experiments to study the determinants of individuals’ and firms’ decisions to pay taxes (Pomeranz and Vila-Belda, 2019; Antinyan and Asatryan, 2024). Prior randomized control trials, typically conducted through messages sent by tax authorities (e.g., letters or emails), have primarily focused on altering taxpayers’ perceptions of the enforcement environment (Slemrod, 2019) or enhancing tax morale by targeting intrinsic motivation, reciprocity, and social norms (Luttmer and Singhal, 2014). We contribute to this line of work by showing that well-targeted messages aimed at improving perceptions about the integrity and quality of tax officials can be highly effective in shifting taxpayer attitudes, translating into increases in tax compliance.

2. Hypotheses

Following standard practice in the political psychology literature, we define beliefs as the initial intuitions and feelings that serve as a baseline for processing new information. These priors are shaped by the accumulation of previous experiences and evaluations, influencing how individuals predispose themselves to accept new information as true or false, favorable or unfavorable.

Priors are often regarded as the “strongest obstacles to political persuasion” (Cotter

et al., 2020a, p. 69). For anti-corruption messages, this presents a unique challenge: such messages may heighten the salience of corruption-related issues without necessarily altering underlying beliefs, particularly when individuals are negatively predisposed to counter-attitudinal information. In contexts where corruption is perceived as pervasive, anti-corruption campaigns may inadvertently reinforce preexisting beliefs, leading to adverse effects rather than fostering change in perceptions or attitudes (Peiffer, 2020; Cheeseman and Peiffer, 2022). For instance, Peiffer and Alvarez (2016) and Peiffer (2020) propose that “positive” messages—those emphasizing government effectiveness in combating corruption—might improve corruption attitudes by introducing new information or persuading citizens of progress. However, their studies reveal that these messages can have the opposite effect, fostering more negative attitudes toward corruption and diminishing perceptions of government effectiveness. As Peiffer (2020) concludes, “if a positively toned message inadvertently primes a negative issue, the message can lead individuals to adopt more negative views about the issue” (p. 7).

Our approach shifts the focus from the tone of anti-corruption messages to their impact on people’s beliefs. Specifically, we evaluate how these messages influence perceptions of corruption, attitudes toward tax evasion, and actual tax behavior. As DellaVigna (2009) note, “persuasion affects behavior because it changes receivers’ beliefs” (p. 17). By examining how belief updates translate into tangible actions, our study sheds light on the mechanisms through which anti-corruption messaging can drive behavioral change, particularly in contexts characterized by deep-seated skepticism toward government institutions. We hypothesize that to effectively influence predispositions and behaviors related to corruption, campaigns must also improve underlying beliefs about the issue, particularly among individuals who are initially predisposed to negative views.

An extensive body of literature underscores how affectively charged priors systematically distort the recognition, comprehension, and evaluation of information, driving both confirmation and disconfirmation biases. These biases make it particularly challenging for counter-attitudinal information to override initial gut-level responses (Taber et al., 2009, p. 66). This prediction of backlash is consistent with several theoretical

mechanisms, including models of motivated reasoning and Bayesian learning (Guess and Coppock, 2020, p. 1497). Despite their theoretical differences, these approaches converge on a common insight: prior beliefs play a central role in shaping how individuals process new evidence. Anticorruption campaigns may simply reinforce perceptions of extensive public corruption.

To counteract entrenched biases, effective messaging must engage with beliefs to increase receptiveness to counter-attitudinal information.² Our experiments implement a straightforward “debiasing” technique aimed at shifting prior beliefs by providing an external benchmark as a reference point to redirect attention away from internal biases (Tversky and Kahneman, 1974). The concept of an external benchmark draws on the idea that individuals often evaluate information using internal reference points, which tend to be biased. For example, perceptions that “the government is corrupt” are typically based on personal expectations of what governance should be. An external benchmark instead contextualizes these beliefs by presenting comparative data—specifically, perceptions of public corruption in Honduras relative to other countries.

Research in political psychology supports the effectiveness of altering reference points to relax entrenched priors (Taber et al., 2009, p. 71). Instead of asking respondents to “take the perspective of others” (e.g., Enríquez et al., 2024) or “keep an open mind” (e.g., Barabas, 2004; Taber and Lodge, 2006), our approach embeds alternative viewpoints directly into the comparative data presented in the campaign messages. By comparing perceptions of corruption in Honduras with those in other Latin American countries, we provide respondents with a credible external reference point, sourced from *Latino-barómetro*, a respected and independent institution unconnected to the government or our research team. This approach reflects findings by Weitz-Shapiro and Winters (2017), who emphasize the importance of credible information sources in shaping citizen responses to corruption-related messages. This design avoids relying on explicit instructions about

²Other studies explore individual-level differences in the propensity for directional reasoning, highlighting factors like political knowledge and the strength of prior attitudes (Taber and Lodge, 2006; Taber et al., 2009). Contextual elements, such as monetary incentives to improve evaluation accuracy (Bullock et al., 2015) and the credibility of information sources (Slothuus and De Vreese, 2010), also influence how individuals process and respond to counter-attitudinal information.

how to process information, offering a more realistic framework for how individuals might encounter and interpret anti-corruption campaign material in everyday contexts.

In crafting our messages, we also draw on the literature on political advertising, which emphasize the effectiveness of “in-group” messengers—such as the “average citizen” and “everyday people” (Hewitt et al., 2024, p. 4)—as particularly powerful sources. The pre-analysis plan proposed the following hypotheses³:

H1: Reform Treatment. Communicating government efforts to combat corruption might have a counterproductive effect, worsening perceptions of the Honduran tax authority and increasing attitudes toward corrupt practices relative to a neutral message (control).

H2: Debiasing Treatment. Presenting information emphasizing that the average Honduran perceives corruption in tax administration to be relatively low compared to citizens in the rest of Latin America will positively influence individuals’ perceptions of the tax authority and reduce willingness to engage in tax evasion compared to a neutral message (control).

H3: Sequential Treatment. The sequential combination of changing beliefs about corruption (“debiasing”) followed by emphasizing government actions (“reform”) could improve more effectively perceptions of the Honduran tax authority and reduce willingness to engage in corrupt behavior compared to a control group.

H4: Sequential Treatment versus Reform Treatment. The “sequential” combination significantly outperforms just emphasizing government actions (“reform”) in terms of reducing willingness to engage in tax evasion. In other words: the same (“reform”) message has a significantly different impact depending on how pessimistic prior beliefs about corruption are.

³Our full pre-analysis plan can be found in Appendix Section D. We have changed the labels of the treatments from the pre-analysis plan as follows: “perception” has been renamed to “debiasing”, “purge” has been renamed to “reform”, and “combined” has been renamed to “sequential”.

3. Case Selection: Tax Administration Reform

Honduras is a lower-middle-income country in Latin America with GDP per capita of \$6,700 PPP in 2022. It ranks 173 out of 213 countries in the World Bank’s Control of Corruption Index. Tax revenues account for 18 percent of GDP and tax collection is heavily reliant on taxes on goods and services, making up over half of the total tax burden.

Starting in 2014, the Honduran tax administration underwent a series of reforms and institutional changes aimed at strengthening the country’s tax system. These included improvements in operational management, a new billing regime, and the adoption of new technologies, among others. The key salient dimension of the reform that we exploit in our experiment was an effort to restore the tax administration’s corporate image through a new human resource policy, motivated in part by the need to address corruption inside the tax administration.

Prior to the reform, the tax administration faced challenges such as insufficient professionalization of its workforce,⁴ lack of transparency and integrity in crucial areas of the organization,⁵ inefficient tax processes, and weak information systems that hampered attempts at evasion control and contributed to poor tax collection performance (e.g., tax revenues hovered at 15 percent of GDP by the time of the reform onset). Citizen perceptions of corruption in the tax administration remained widespread: more than 40 percent of Honduran citizens believed that all or almost all tax officials were involved in corrupt practices (Latinobarómetro 2016).

Recognizing the need for a comprehensive overhaul, between 2014 and 2019 the government implemented a major institutional reform of the tax administration, with support from multilateral organizations. The reform involved the dissolution of the existing tax administration by executive decree, including the dismissal of 1,500 employees, consti-

⁴For instance, only a third of the tax administration personnel held a college degree by the time of the reform onset. See IDB (2015).

⁵Prior to the reform, out of a sample of more than 800 employees who took a polygraph test, 30 percent failed it. The tax administration received, on average, 90 complaints per month related to issues of lack of transparency in job performance. See IDB (2015).

tuting 85 percent of existing personnel.⁶ Approximately 300 employees were retained to safeguard the tax base during the transition period. The recruitment and selection process of personnel was conducted under the guidelines of a new human resource policy, including: (i) a tax career path with respective job profiles, (ii) competitive salary scales and performance evaluation procedures, and (iii) promotion schemes. New hires were subject to several tests prior to selection, including polygraph evaluations. By January 2017, the new tax administration, *Servicio de Administración de Rentas* (SAR), started operations with around 500 workers, a combination of employees from the transition period and new hires.

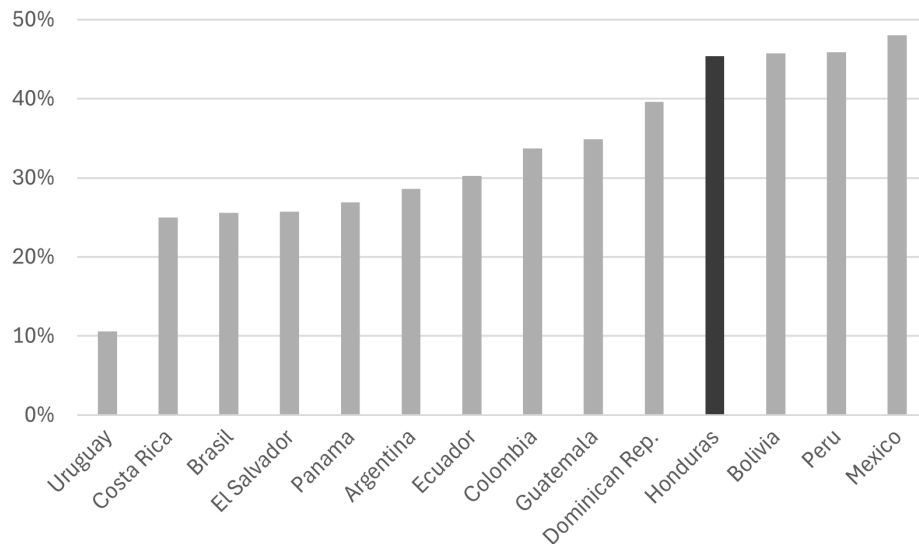
The tax administration reform significantly expanded the tax base, tripling the number of active taxpayers. It also reduced compliance costs by increasing electronic filings of major taxes from less than 50% to over 95%, and enhanced the quality of human resources within the tax administration, with the share of personnel holding college degrees more than doubling. These reforms played a crucial role in improving tax collection, which remained 20% above pre-reform levels until the onset of the COVID-19 crisis. Additionally, public perceptions of corruption within the tax administration have markedly improved. According to nationally representative surveys, the Honduran tax authority is now perceived as one of the least corrupt in Latin America (Latinobarómetro 2020).

Figure 1 illustrates Honduras's remarkable progress in the regional tax administration corruption ranking. In particular, Honduras moved from 11th to 2nd place among countries perceived to have the least corruption in their tax administration. We leverage this progress, alongside government actions to combat corruption within the tax authority, to evaluate their effectiveness in changing respondents' beliefs, attitudes toward the tax authority, and willingness to pay taxes.

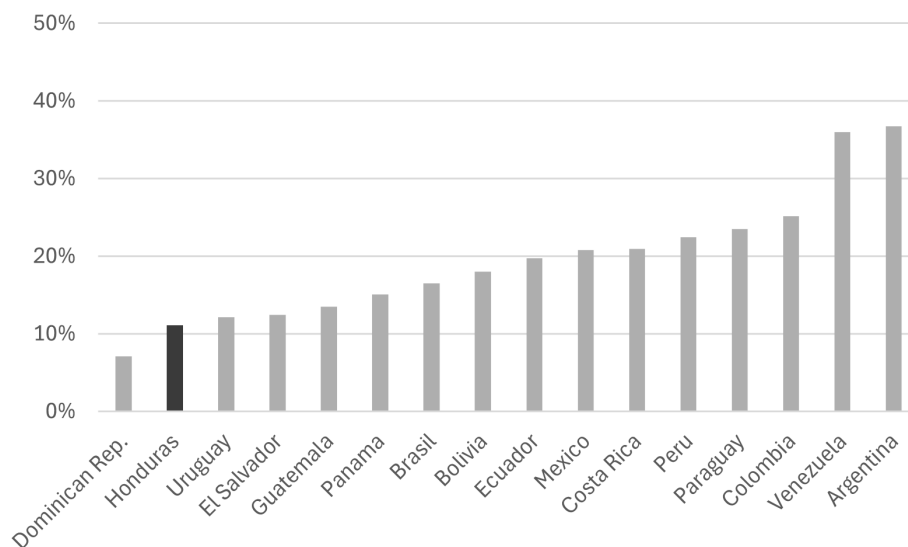
⁶A special commission was appointed to proceed with termination of existing contracts. Additionally, 900 employees from the customs area were transferred to the Ministry of Finance.

Figure 1: Corruption Perceptions of Tax Administration - Latinobarometro

(a) Latinobarometro 2016



(b) Latinobarometro 2020



Notes: The figure shows the answer to the question “How many of the following people do you think are involved in corruption, or haven’t you heard enough about them to say?” for the years 2016 and 2020 of the Latinobarometro survey, showing the response for the tax administration. Each panel shows the proportion of people who believe that the tax administration is corrupt by country, these being ordered from lowest perceptions of corruption to highest. In 2016, individuals who believed corruption existed were identified as those who reported that “almost all” or “all” officials were corrupt, whereas in 2020, they could only specify whether the tax administration was involved in corruption.

4. The Survey Experiment

To assess how individuals react to anticorruption efforts by the tax administration, we designed a survey experiment with data collected through an online survey between October 2, 2023, and October 4, 2023. We recruited 2,372 participants through Facebook ads (2,000 targeted), but excluded from our main sample those individuals who did not pass our attention check, who had their IP address duplicated within the sample, or who had their IP address duplicated with our pilots, resulting in our final sample of 1,411 observations. Participants were randomly selected from the general population of Honduras (with internet access) and they were selected in such a way as to match the census data of Honduras in terms of gender and by region.

4.1. Treatments

Respondents in our survey were assigned to one of four information treatments:⁷

1. The control group, which received basic information about the SAR's role in tax collection, along with an image of the locations of the SAR offices in Honduras.
2. The “reform” treatment arm, which transmits the SAR's recognition of corruption as a significant issue and recent efforts to combat it. The message highlights the personnel renewal done by the government for the purpose of reducing corruption within the tax agency and includes a simple figure to illustrate the magnitude of the policy reform.
3. The “debiasing” treatment arm, which provides basic information about the SAR's role in tax collection and information about the SAR's reputation as one of the least corrupt tax authorities in Latin America, backed by a graph using survey data indicating it as the second most honest tax authority institution in the Latin American region.

⁷See Appendix Section B for the full survey instrument and treatment layouts.

4. The “sequential” treatment arm, which combines the two previous information treatments, with the “debiasing” message shown first, followed by the “reform” treatment.

Participants were randomly assigned to each group with equal probability (25 percent to each group). They were also informed about the anonymity of their answers. Appendix Table A1 presents a balance check among the treatment groups based on our pre-treatment variables. Appendix Table A2 presents the descriptive statistics for our outcomes and main pre-treatment variables (used in the heterogeneity analysis).

4.2. Outcomes

The primary focus of this analysis is respondents’ beliefs and perceptions regarding the tax authority and pervasive corruption. Respondents rated their beliefs on a scale from 0 (not likely at all) to 10 (very likely) on the likelihood of SAR officials engaging in corruption (*Perception of Corruption of the SAR*) and the chance that a SAR official might seek gifts or money for resolving or preventing tax-related issues (*Probability That a SAR Agent Will Accept a Bribe*). Aligning with standard practices for multiple outcome analysis, we developed the *Beliefs Index* combining these two outcomes. Since both outcomes are measured on the same scale, this index is simply the average of both outcomes.

Further, we examine secondary outcomes concerning individuals’ willingness to evade taxes. To counter potential biases in self-reported measures of tax evasion, we employed a crosswise measure. This involved respondents reporting the truthfulness of two statements: one on their willingness to evade taxes and the other, a non-sensitive question about the birth month of their mother. With the known probability distribution of the latter, unbiased prevalence estimates on the stated willingness to evade taxes were obtained, following Jann et al. (2012). We alter our outcome in the following way:

$$\tilde{R} = (R + p - 1)/(2p - 1) \tag{1}$$

Where p is the probability that an individual’s mother was born in November or Decem-

ber, which on average is $p = 2/12$, and R is a dummy variable equal to 1 if individuals have responded that both statements (the one about their willingness to evade taxes and the non-sensitive statement) are true or that both statements are false, while the dummy is equal to 0 if respondents reported only one of the statements is true.⁸

We also evaluated respondents’ reactions to a hypothetical scenario where a store owner offers a VAT-exempt discount, probing their readiness to participate in tax evasion. Respondents were presented with a scenario in which a store owner offered a 10 percent discount in exchange for a sale without VAT during a shopping transaction. Respondents indicated whether they were willing to accept this discount, implying a willingness to evade the corresponding tax.

Additionally, we incorporated a series of questions to probe our belief-updating mechanism. Specifically, we asked them to evaluate the *SAR’s Capability*—their assessment of the tax administration’s ability to collect taxes owed by taxpayers. To examine broader effects beyond taxation, we included two additional measures: *Corruption of City Residents*, which captured perceptions of fellow citizens’ involvement in corrupt activities, and a behavioral test, *Lies in Dice Game*. In this game, participants had the opportunity to misreport their dice rolls to improve their lottery odds, serving as a proxy for their propensity to provide dishonest information. These outcomes allowed us to assess whether the messages influenced participants’ perceptions of corruption and dishonest behavior in general, extending beyond the specific context of taxation.⁹

4.3. Estimation

We estimate the following linear regression model by Ordinary Least Squares for each of our outcomes:

$$Y_i = \alpha + \beta_1 \text{Reform}_i + \beta_2 \text{Debiasing}_i + \beta_3 \text{Sequential}_i + \beta_4 X_i + \varepsilon_i \quad (2)$$

The specification includes indicator variables for each treatment arm with the control

⁸For more information about this transformation see Jann et al. (2012).

⁹These outcomes were pre-registered as secondary hypotheses.

group as the reference category. Y_i is the outcome analyzed, while X_i is a control vector chosen by running a double-selection lasso linear regression for each outcome, following the recommendations of Cilliers et al. (2023).¹⁰ ε_i is the error term, estimated using Huber-White standard errors to account for the potential heteroskedasticity. We report the coefficients β_1 , β_2 , and β_3 for each of our regressions, while also including the mean of the control group as a comparison. Finally, we also include for each outcome two different tests in which we compare whether the coefficient of the “sequential” treatment is the same as that of the “debiasing” or “reform” treatment, respectively.¹¹

4.4. Main Results

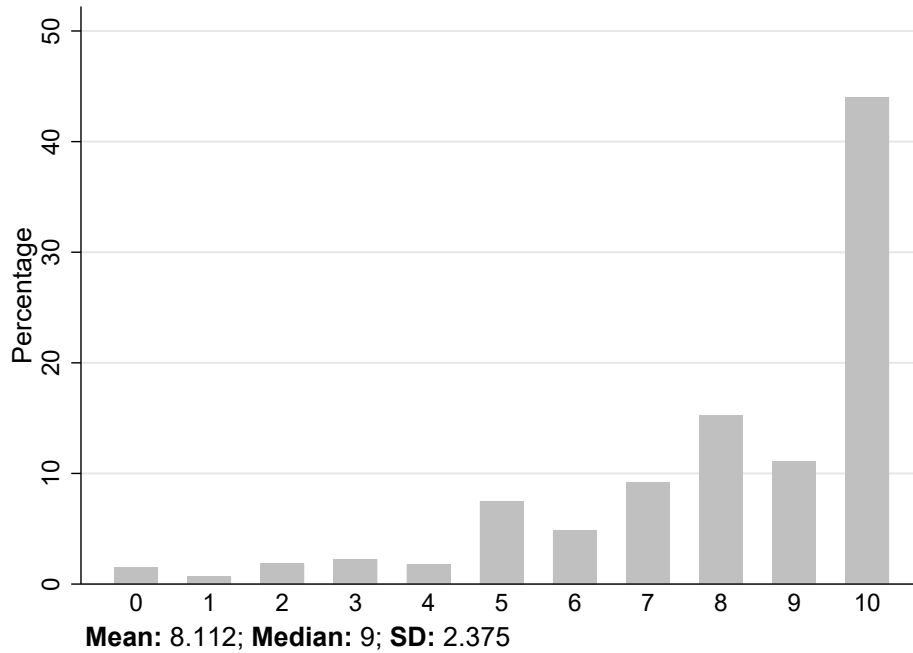
We first document extremely pessimistic beliefs about generalized corruption. As depicted in the Histogram of Pre-Treatment Corruption Beliefs (Figure 2), around 70% of respondents assign an 8 or more to the likelihood that public administration officials engage in corrupt practices, with nearly 45% of the sample assigning the maximum likelihood. These beliefs reveal widespread skepticism about public-sector integrity, establishing an important baseline for assessing whether our interventions can shift entrenched perceptions.

We then assess the effect of SAR’s messages on our primary outcomes, beginning with the *Beliefs Index*. As displayed in columns (1) and (4) of Table 1, the impact of the treatments on individual beliefs is nuanced. The “reform” treatment showed a positive but statistically insignificant effect on perceptions of corruption. Conversely, the “debiasing” treatment—which highlighted SAR’s standing as one of the least corrupt tax agencies in the region—produced a statistically significant reduction in corruption beliefs, lowering the *Beliefs Index* by an average of 0.43 units when controls are included (column (4)). With a baseline belief of 6.2 units in the control group, this effect translates

¹⁰The potential controls, ultimately selected or excluded by LASSO, included respondent characteristics such as sex, aggregated age groups, ideological self-placement, education level, confidence in institutions (e.g., church, congress, executive power, and police), and prior perceptions of corruption among administrative officials and Hondurans in general.), and region.

¹¹Appendix Table A3 tests for potential experimenter demand effects by examining whether treatments influence two outcomes: the survey rating (number of stars) and the probability of leaving a comment different than “none” or similar. Results show no significant effects of the treatments on either outcome, indicating that experimenter demand effects are unlikely.

Figure 2: Histogram of Pre-Treatment Corruption Beliefs of the Public Administration



Notes: The figure shows a histogram of the responses to the question “On a scale from 0 to 10, where 0 is “not likely at all” and 10 is “very likely”, how likely do you think public officials in Honduras are involved in acts of corruption?” The mean, median and standard deviation of the variable are also included.

to a 7 percent decrease in perceived corruption. The most pronounced impact came from the “sequential” treatment, which combined the “debiasing” and “reform” messages. This approach resulted in a significant reduction in the *Beliefs Index*, decreasing it by an average of 0.78 units with controls, representing a notable 13 percent decline in corruption beliefs.

The last two rows show significant differences between the sequential treatment and other treatments. The row that tests the “reform” effect against the “sequential” one shows a large and significant difference, which confirms our hypothesis that changing individuals’ initial corruption perceptions leads to improvements in the reception of anti-corruption messages. We also show that the difference between the “sequential” and “debiasing” treatment arm is statistically significant if we include controls, however only at a 10% significance level. Appendix Table A4 present these results for each component of the *Beliefs Index* separately.

Table 1 also examines the *Willingness to Evade Taxes* outcome (columns (2) and (5)). The coefficient associated with the “reform” message is positive but not statistically

Table 1: Beliefs Index and Willingness to Pay Outcomes - Survey Experiment

Variables	No controls			Controls added		
	(1) Beliefs Index	(2) Willing to Evade Taxes	(3) Accepts Discount	(4) Beliefs Index	(5) Willing to Evade Taxes	(6) Accepts Discount
Reform	0.085 (0.201)	1.721 (5.616)	1.666 (3.552)	0.263 (0.169)	2.173 (5.572)	2.680 (3.440)
Debiasing	-0.699*** (0.209)	-7.330 (5.654)	-2.148 (3.525)	-0.433** (0.182)	-7.027 (5.644)	-0.604 (3.397)
Sequential	-1.025*** (0.200)	-11.631** (5.478)	-4.010 (3.397)	-0.778*** (0.179)	-11.159** (5.449)	-2.642 (3.267)
Observation	1411	1411	1411	1411	1411	1411
Control Group Mean	6.150	51.4%	33.3%	6.150	51.4%	33.3%
LASSO controls	No	No	No	Yes	Yes	Yes
Test Reform = Sequential	0.000	0.018	0.107	0.000	0.017	0.125
Test Debiasing = Sequential	0.122	0.449	0.594	0.068	0.467	0.556

Notes: Robust standard errors in parentheses. “Beliefs Index” outcome is the mean between the *Belief of Corruption of the SAR* outcome and the *Probability that a SAR Agent will Accept a Bribe* outcome, which varie from 0 (not likely at all) to 10 (very likely). “Willing to Evade Taxes” outcome is a dummy variable, which was transformed according to Jann et al. (2012) in order to get the unbiased estimates of the willingness of evading taxes from the crosswise initial measure. Both outcomes varie from 0 (not likely at all) to 10 (very likely). “Accepts Discount” outcome equal to 1 if the respondent is willing to accept buying without VAT included. Controls were chosen using a double-selection lasso linear regression from the following possibilities: respondent’s sex, aggregated age groups, ideological self-placement, education level, confidence in institutions (church, congress, executive power, police), prior perceptions of corruption (administrative officials and Hondurans in general), and region. Dummy variables were multiplied by 100 so that the interpretation of the coefficients in percentage points is direct. The last two rows present the p-values for two tests of coefficient equality: the first compares the coefficient for the “reform” treatment with that for the “sequential” treatment, while the second compares the coefficient for the “debiasing” treatment with the “sequential” treatment. P-values: ***p<0.01, **p<0.05, *p<0.1.

significant at conventional levels. In turn, the “debiasing” treatment consistently had a negative effect on people’s willingness to evade taxes; however, again, none of the coefficients are statistically significant at the conventional level. Combining both messages, however, had a substantial, statistically and economically significant impact on the individuals’ willingness to evade taxes, reducing it by approximately 11 percentage points on average (column (5)). Given a prevalence estimate of 51 percent in the control group, the treatment effect is able to reduce the share of evaders by around 21 percent. Consistent with the results on beliefs, the difference between treatments is significant when comparing the “sequential” and “reform” treatment arms at a 5% significance level.

These findings demonstrate that the “sequential” treatment effectively shifts both corruption beliefs and attitudes toward tax evasion. As the last row of Table 1 shows,

similar to its effect on beliefs, the communication of a specific government anti-corruption policy has a statistically differential effect if we first change the initial corruption beliefs of the respondents. This overcomes the backfiring effect identified in previous studies and provides room for optimism, challenging earlier findings that led scholars to argue that “policy makers and activists should avoid explicitly invoking the government and the state in their campaigns” (Cheeseman and Peiffer, 2023, p. 1047), while aligning with more recent findings (Erllich and Gans-Morse, 2025).

Not all evidence, however, aligns with our hypotheses. We also included a question assessing respondents’ likelihood of colluding with a seller to avoid paying sales tax. The analysis of the *Accepts Discount* question, presented in columns (3) and (6) of Table 1, reveals no significant average effects from our treatments on respondents’ willingness to accept a VAT-free discount. This null result may stem from the nature of the outcome itself: self-reported behaviors are prone to social desirability bias, where respondents report what they perceive as morally acceptable rather than their actual behaviors. Nevertheless, the field experiment analyzed below demonstrates that our messages did influence real-world behavior, leading to substantial improvements in tax compliance.

4.5. Further analysis

In the previous sections, we argued that providing a benchmark to contextualize beliefs—showing subjects that Hondurans, on average, hold more favorable views of their tax agency compared to other Latin American countries—could improve priors and encourage more positive tax behaviors. To explore this mechanism further, we pre-registered additional hypotheses to examine the interplay between corruption-prevention messaging, citizens’ beliefs about government integrity, and their propensity for tax evasion. Three key pieces of evidence support our proposed belief-update mechanism.

First, our messages effectively influenced beliefs about SAR’s enforcement capabilities. Column (4) of Table 2 reveals that all treatments improved perceptions of SAR’s enforcement capacity, with the “sequential” treatment showing the largest effect, increasing perceptions by 0.23 units (7 percent). While the “reform” treatment had the weakest

Table 2: Beliefs About SAR’s Capability, Perceptions of City’s Corruption, and Dishonest Behavior Outcomes

Variables	No controls			Controls added		
	(1) SAR’s Capability	(2) Corruption of city’s residents	(3) Lies in Dice Game	(4) SAR’s Capability	(5) Corruption of city’s residents	(6) Lies in Dice Game
Reform	0.171* (0.088)	0.128 (0.207)	3.983* (2.297)	0.154* (0.086)	0.179 (0.176)	3.957* (2.182)
Debiasing	0.208** (0.091)	0.227 (0.203)	1.618 (2.194)	0.177** (0.088)	0.248 (0.182)	1.226 (2.143)
Sequential	0.276*** (0.087)	0.256 (0.202)	0.846 (2.087)	0.230*** (0.086)	0.200 (0.173)	0.098 (2.041)
Observation	1411	1411	1411	1411	1411	1411
Control Group Mean	3.143	5.520	8.4%	3.143	5.520	8.4%
LASSO controls	No	No	No	Yes	Yes	Yes
Test Reform = Sequential	0.242	0.532	0.183	0.384	0.903	0.088
Test Debiasing = Sequential	0.462	0.885	0.732	0.558	0.784	0.605

Notes: Robust standard errors in parentheses. “Corruption of city’s residents” outcome is the respondent belief that individuals of his/her city are prone to corruption, which varies from 0 (not likely at all) to 10 (very likely). “SAR’s Capability” outcome corresponds to the respondents’ belief that the SAR has the capability of collecting the amounts owed by individuals with tax obligations, which varies from 1 (not capable at all) to 5 (highly capable). “Lies in Dice Game” variable equal to 1 if individual lied on the dice game. Controls were chosen using a double-selection lasso linear regression from the following possibilities: respondent’s sex, aggregated age groups, ideological self-placement, education level, confidence in institutions (church, congress, executive power, police), prior perceptions of corruption (administrative officials and Hondurans in general), and region. Dummy variables were multiplied by 100 so that the interpretation of the coefficients in percentage points is direct. The last two rows present the p-values for two tests of coefficient equality: the first compares the coefficient for the “reform” treatment with that for the “sequential” treatment, while the second compares the coefficient for the “debiasing” treatment with the “sequential” treatment. P-values: ***p<0.01, **p<0.05, *p<0.1.

effect (only significant at a 10% level), differences between treatments were not statistically significant. This finding is consistent with our prior results that the “debiasing” and “sequential” treatments reduced perceptions of corruption within SAR and the likelihood of SAR agents accepting bribes.

In contrast, column (5) shows no treatment effects on perceptions of corruption among city residents, regardless of the treatment arm. Taken together, these results suggest that given that our intervention altered tax compliance willingness without changing beliefs about the propensity of other citizens to evade taxes, we can infer that the positive update in beliefs is primarily related to a shift in the perceived competence or efficacy of the SAR in detecting tax evasion.

We also assessed the impact of the messages on honesty-related behaviors using a

dice game, where participants could misreport their dice rolls to increase their chances of winning lottery tickets. The virtual die was biased toward 1, incentivizing dishonest reporting. Column (6) reveals no statistically significant effects on dishonest reporting, except for the “reform” treatment, which backfired, increasing dishonesty by approximately 4 percentage points compared to the 8 percent baseline in the control group, and only statistically significant at a 10% significance level.¹² These findings suggest that while targeted interventions can effectively shift beliefs and behaviors directly related to the intervention, they may leave broader considerations, such as unrelated dishonest behavior, unchanged.

Second, we demonstrate that the effects of the debiasing and sequential messages were particularly pronounced among individuals with more negative priors. To explore this issue, we conducted a heterogeneity analysis by estimating interactions between the treatments and a dummy variable indicating whether respondents’ pre-treatment beliefs about corruption exceeded the median.¹³ This approach enables formal testing of heterogeneity hypotheses. Specifically, we use the following linear regression model, estimated using Ordinary Least Squares:

$$\begin{aligned}
Y_i = & \alpha + \beta_1 \text{Reform}_i + \beta_2 \text{Debiasing}_i + \beta_3 \text{Sequential}_i + \beta_4 \text{Prior}_i^+ \\
& + \beta_5 \text{Prior}_i^+ \cdot \text{Reform}_i + \beta_6 \text{Prior}_i^+ \cdot \text{Debiasing}_i + \beta_7 \text{Prior}_i^+ \cdot \text{Sequential}_i \\
& + \beta_8 X_i + \varepsilon_i
\end{aligned} \tag{3}$$

This equation builds on equation (2) by incorporating the variable Prior_i^+ , a dummy variable indicating whether the individual’s pre-treatment corruption belief regarding public administration officials exceeded the median. All other variables remain consistent with equation (2). As shown in Figure 2, the median pre-treatment corruption belief was 9. This analysis, therefore, focuses on testing whether treatment effects differ depending

¹²Differences across treatments were not statistically significant at conventional levels.

¹³The question asked was: “On a scale from 0 to 10 , where 0 is “not likely at all” and 10 is “very likely”, how likely do you think public officials in Honduras are involved in acts of corruption?”

on whether individuals had extremely negative or moderately negative corruption beliefs prior to the intervention.

Table 3 presents the results of the heterogeneity analysis for the *Beliefs Index* and *Willingness to Evade Taxes* outcomes. For clarity in presentation, this table reports only the linear combinations of the interaction terms for each group, facilitating interpretation. Among individuals with lower pre-treatment perceptions of public corruption (below the median), the “reform” treatment significantly increased corruption perceptions by 0.44 points when controls were included (column (3)). However, none of the other treatments yielded statistically significant results.¹⁴ For the *Willingness to Evade Taxes* outcome, shown in column (4), the observed patterns align with our expectations, but none of the treatments resulted in statistically significant effects.

In contrast, for individuals with higher pre-treatment corruption beliefs (above the median), the effects were stronger, and the sequential treatment produced statistically significant results across both outcome variables. The “debiasing” treatment significantly reduced corruption perceptions, with a decrease of 0.68 points when controls were included. The “sequential” treatment showed the largest impact, reducing the *Beliefs Index* by 1.52 points on average, equivalent to a 25 percent decrease in perceived corruption relative to the control group mean. These findings suggest that combining treatments amplifies their impact on beliefs compared to delivering them individually. Furthermore, both the “reform” and “sequential” treatments demonstrated statistically significant differences between belief groups, highlighting that the “reform” message alone was less effective in reducing corruption perceptions. Regarding the *Willingness to Evade Taxes*, the sequential treatment resulted in the most substantial reduction, decreasing tax evasion willingness by 14.3 percentage points when controls are included (see column (4)).

Finally, to further assess the hypothesis that changes in individuals’ willingness to pay taxes occur through changes in their beliefs about the tax administration’s corruption standing, we employ an instrumental variables approach. We use the three different

¹⁴The sequential treatment produced a statistically significant reduction (p-value < 0.1) in corruption perceptions without controls, but this effect dissipates when controls are included.

Table 3: Heterogeneity Analysis of Beliefs Index and Willingness to Evade Outcomes

Variables	No Controls		Controls Added	
	(1) Beliefs Index	(2) Willing to Evade Taxes	(3) Beliefs Index	(4) Willing to Evade Taxes
<i>Low Pre-Treatment Corruption</i>				
Reform - Low Corruption	0.152 (0.248)	9.468 (7.699)	0.444** (0.222)	11.095 (7.690)
Debiasing - Low Corruption	-0.411 (0.257)	-4.813 (7.578)	-0.205 (0.238)	-3.436 (7.649)
Sequential - Low Corruption	-0.403* (0.242)	-7.466 (7.252)	-0.244 (0.218)	-6.189 (7.280)
Test Debiasing = Sequential - Low Corrup.	0.976	0.720	0.864	0.714
Test Reform = Sequential - Low Corrup.	0.019	0.025	0.001	0.022
<i>High Pre-Treatment Corruption</i>				
Reform - High Corruption	-0.046 (0.273)	-7.205 (8.203)	0.065 (0.262)	-6.187 (8.266)
Debiasing - High Corruption	-0.913*** (0.300)	-9.739 (8.488)	-0.676** (0.285)	-10.35 (8.548)
Sequential - High Corruption	-1.594*** (0.311)	-15.90* (8.399)	-1.524*** (0.303)	-14.35* (8.459)
Test Debiasing = Sequential - High Corrup.	0.000	0.312	0.000	0.347
Test Reform = Sequential - High Corrup.	0.049	0.487	0.010	0.660
<i>Tests Low Corruption=High Corruption Coefficients</i>				
Reform Low = Reform High	0.592	0.139	0.267	0.126
Debiasing Low = Debiasing High	0.204	0.665	0.201	0.547
Sequential Low = Sequential High	0.003	0.447	0.001	0.466
Observation	1411	1411	1411	1411
LASSO controls	No	No	Yes	Yes

Notes: Robust standard errors in parentheses. “Beliefs Index” outcome is the mean between the *Belief of Corruption of the SAR* outcome and the *Probability that a SAR Agent will Accept a Bribe* outcome, which varie from 0 (not likely at all) to 10 (very likely). “Willing to Evade Taxes” outcome is a dummy variable, which was transformed according to Jann et al. (2012) in order to get the unbiased estimates of the willingness of evading taxes from the crosswise initial measure. Dummy variables were multiplied by 100 so that the interpretation of the coefficients in percentage points is direct. Controls were chosen using a double-selection lasso linear regression from the following possibilities: respondent’s sex, aggregated age groups, ideological self-placement, education level, confidence in institutions (church, congress, executive power, police), prior perceptions of corruption (administrative officials and Hondurans in general), and region. Section *Low Pre-Treatment Corruption* of the table shows the effects of our treatments for those individuals below the median pre-treatment corruption beliefs (the distribution of these beliefs can be seen in Figure 2), while also including tests that compare the equality of our treatments. Section *High Pre-Treatment Corruption* of the table shows the effects of our treatments for those individuals above the median pre-treatment corruption beliefs, while also including tests that compare the equality of our treatments. Finally, section *Tests Low Corruption=High Corruption Coefficients* compares the equality of the coefficients for each treatment between those individuals above and below the median pre-treatment corruption beliefs. P-values: ***p<0.01, **p<0.05, *p<0.1.

treatments as instruments for the *Beliefs Index* variable in the first stage.¹⁵ As discussed by Haaland et al. (2023), if the treatment affects certain beliefs without influencing others—which is reasonable in our context, as Table 2 shows no impact on, for example, perceptions of corruption in the respondent’s city—then it is possible to isolate the effect of beliefs on the willingness to evade taxes.

Table 4 indicates that corruption beliefs significantly influence the willingness to evade taxes, with an increase of one unit in corruption beliefs raising the probability of tax evasion by around 13 percentage points on average. At the same time, government corruption beliefs do not significantly impact the probability of accepting a discount or lying in the dice game, consistent with our findings that, while corruption beliefs were altered (as shown in column (4) of Table 1), the treatments did not have a significant impact on average on these willingness-to-pay outcomes (as shown in column (6) of Table 1 and in column (6) of Table 2). The effect was significant only for respondents assigned to the reform treatment in the probability of lying in the dice game, and only at a 10% significance level.¹⁶

5. Field Experiment

Even though our survey experiment mitigates social desirability bias through the self-reported willingness-to-pay measure, it cannot answer whether changes in beliefs and willingness to pay affect actual tax behavior. To address external validity concerns, we analyze results from a field experiment conducted by the Honduran Tax Administration (SAR) before the 2023 income tax filing deadline. Building on our survey findings, SAR randomly assigned similar treatment messages to individuals and firms subject to the income tax to test whether our experimental results would translate into tax compliance behavior.

SAR sent emails to a representative sample of 45,000 taxpayers in two rounds: three

¹⁵For similar applications, see Bottan and Perez-Truglia (2020); Roth and Wohlfart (2020); Cullen and Perez-Truglia (2022).

¹⁶Regarding the possibility of a weak instrument, our F-statistic exceeds the usual rule of thumb of 10 in all cases.

Table 4: Instrumental Variables Estimations

Variables	(1) Willing to Evade Taxes	(2) Accepts Discount	(3) Lies in Dice Game
<i>Second Stage Estimates</i>			
Beliefs Index	12.946** (5.333)	4.214 (2.988)	2.871 (1.975)
<i>First Stage Estimates</i>			
Reform	0.271 (0.172)	0.282 (0.173)	0.266 (0.173)
Debiasing	-0.449** (0.187)	-0.448** (0.187)	-0.449** (0.187)
Sequential	-0.798*** (0.183)	-0.790*** (0.184)	-0.797*** (0.183)
Observations	1411	1411	1411
LASSO controls	Yes	Yes	Yes
F-stat	13.14	13.10	13.02

Notes: Robust standard errors in parentheses. ‘Willing to Evade Taxes’ outcome is a dummy variable, which was transformed according to Jann et al. (2012) in order to get the unbiased estimates of the willingness of evading taxes from the crosswise initial measure. ‘Accepts Discount’ outcome equal to 1 if the respondent is willing to accept buying without VAT included. ‘Lies in Dice Game’ variable equal to 1 if individual lied on the dice game. Dummy variables were multiplied by 100 so that the interpretation of the coefficients in percentage points is direct. F-stat row reports the Kleibergen-Paap rk Wald F statistic. All regressions include controls, which were chosen using a double-selection lasso linear regression from the following possibilities: respondent’s sex, aggregated age groups, ideological self-placement, education level, confidence in institutions (church, congress, executive power, police), prior perceptions of corruption (administrative officials and Hondurans in general), and region. P-values: ***p<0.01, **p<0.05, *p<0.1.

weeks and one week prior to the income tax filing deadline for fiscal year 2023 (April 30, 2024). All emails included a reminder about the filing deadline. Although SAR initially contacted 45,000 firms or individuals, only around 33,000 taxpayers submitted their tax returns during this round.

Although the government did not pre-register this experiment, and we did not participate in its execution, the evidence we analyzed from administrative records—kindly provided by the tax administration—supports the random assignment of the treatment (see Appendix Table A5). This allows us to analyze the data similarly to how researchers ap-

proach natural experiments with observational data (Dunning, 2012). The model emails used by the tax administration are available in the Appendix Section C.

5.1. Treatments

The control group received an email with information about the locations of SAR offices across Honduras, while taxpayers in the treatment group received either the “de-biasing” message or the “sequential” message. This experiment deliberately excluded the “reform” treatment due to its potential to backfire—a concern highlighted by our initial survey results. The differences in the content of messages between our survey experiment and SAR’s intervention are subtle and can be found in the Appendix Section C.¹⁷

5.2. Outcomes

The primary outcome of this analysis is the declared income reported by taxpayers to the tax authority for the 2023 period, measured in thousands of lempiras. This variable directly captures individuals’ tax behavior, enabling an assessment of the interventions’ impact on tax compliance. In 2023, taxpayers in our sample reported a median income of 418,795 lempiras (37,814 USD PPP). For comparison, the 2021 median declared income was 310,691 lempiras (29,338 USD PPP).

5.3. Estimation

Similarly to the approach taken in Equation (2), we estimate the following linear regression model using Ordinary Least Squares (OLS) to assess the effects of the experimental intervention on tax compliance:

$$Y_i = \alpha + \beta_1 \text{Sequential}_i + \beta_2 \text{Debiasing}_i + \beta_3 X_i + \varepsilon_i \quad (4)$$

Here, Y_i represents the declared income, Sequential_i and Debiasing_i are the two treat-

¹⁷In SAR’s experiment, neither group received basic information about SAR’s role in tax collection; the “de-biasing” treatment used a relatively shorter message, and the “sequential” treatment did not include details about the process used to select new employees (e.g., lie detector tests).

ments implemented by the tax authority, and X_i denotes a control vector selected through a double-selection lasso regression applied to the outcome.¹⁸ The error term, ε_i , is estimated using Huber-White standard errors to account for potential heteroskedasticity. We report the coefficients β_1 and β_2 for each regression and include the mean of the control group as a reference point, along with the p-values to test whether the coefficient for the “sequential treatment” differs from that of the “debiasing” treatment.

We estimate Equation (4) on two samples: the full sample of taxpayers and a more limited sample of taxpayers classified with higher risk of non-compliance.¹⁹

5.4. Main Results

As shown in Table 5, the “sequential” treatment proved to be effective, resulting in an increase of 180 thousand lempiras in the reported income (p-value < 0.1), representing a 7 percent increase relative to the control group, when including LASSO controls. The impact is more pronounced for the subsample including only high-risk taxpayers, with an increase of 875 thousand lempiras in taxable income, as shown in column (4). On the other hand, the “debiasing” treatment shows no statistically significant impact across specifications. The p-values testing equality between the effects of the “debiasing” and “sequential” treatments suggest that the differences are significant in all cases (although only at a 10% significance level). These findings align with our hypothesis that the “sequential” treatment reinforces the “debiasing” treatment without negatively influencing it through a priming or confirmation bias effect, as happened with the “reform” treatment in the survey experiment (Table 2).

Overall, the average treatment effects of the communication intervention on actual tax behavior are consistent with results stemming from our self-reported willingness to

¹⁸The potential controls, ultimately selected or excluded by LASSO, encompassed taxpayers’ baseline characteristics, including their region, their risk of non-compliance level, their type (natural or legal entity), the source of information for inclusion in the sample (e.g., previous tax returns, third-party information), and their size.

¹⁹The risk classification is derived from SAR’s risk model, which assesses taxpayers’ non-compliance risks across various compliance dimensions: registration, filing, payment, and accuracy. This model integrates data from tax returns, third-party information, and other government agencies to generate a summary risk score by combining variables for probability (likelihood of risk) and impact (economic damage caused by non-compliance).

Table 5: WTP Outcome - Field Experiment

Variables	No controls		Controls added	
	(1) Reported Income	(2) Reported Income	(3) Reported Income	(4) Reported Income
Debiasing	10.95 (112.55)	165.92 (479.72)	-33.48 (99.93)	3.00 (423.44)
Sequential	312.21*** (118.96)	1,218.31** (500.24)	180.43* (104.60)	874.58** (443.95)
Observations	32,938	6,142	32,938	6,142
Control Group Mean	2,639	7,614	2,639	7,614
LASSO controls	No	No	Yes	Yes
Low-Risk of Payment Excluded	No	Yes	No	Yes
Test Debiasing = Sequential	0.027	0.067	0.074	0.086

Notes: Robust standard errors in parentheses. Observations above the 99th percentile of the declared income were excluded from the regressions. “Reported Income” is the amount of declared income subject to the income tax corresponding to the 2023 period, divided by 1,000. Columns (2) and (4) exclude low risk taxpayers. Controls were chosen using a double-selection lasso linear regression from the following possibilities: region, taxpayer risk level, type of taxpayer (natural or legal entity), the source of information for inclusion in the sample (e.g., previous tax returns, third-party information), and taxpayer size. The last row present the p-value for the test of coefficient equality between the “debiasing” treatment with that for the “sequential” treatment. P-values: ***p<0.01, **p<0.05, *p<0.1.

pay outcome of Table 1, thus reinforcing the validity of our survey experiment measures as proxies for the willingness to evade taxes in the real world.

6. Conclusions and Policy Implications

Using evidence from both a survey experiment and a field experiment conducted in Honduras, this study evaluates the effectiveness of different anticorruption messages on corruption perceptions and willingness to pay taxes. Prior research highlights significant challenges in communicating anticorruption efforts effectively, given mounting evidence of affective biases that influence how individuals process information.

Several studies document a “backfire” effect in which government campaigns to combat corruption inadvertently reinforce negative perceptions. For instance, Peiffer (2020) found that in Jakarta, Indonesia, both “negative” messages emphasizing corruption preva-

lence and “positive” messages highlighting government successes yielded similarly negative effects on perceptions of corruption. Likewise, Cheeseman and Peiffer (2022) reported that anticorruption messages in Lagos, Nigeria, often failed to deter bribery and, in some cases, increased willingness to pay bribes, with outcomes strongly influenced by individuals’ preexisting beliefs about corruption prevalence. However, evidence from a series of experiments in Ukraine suggests that, when messages are perceived as credible and framed appropriately, norm-based anticorruption campaigns can reduce bribery intentions without backfiring (Erlach and Gans-Morse, 2025).

Within the Latin American context, studies also reveal mixed impacts of anticorruption messages. For example, Beesley and Hawkins (2022) found that informing citizens about the positive and negative consequences of corruption reduced trust in political institutions. Similarly, Corbacho et al. (2016) demonstrated that increasing beliefs about corruption prevalence heightened participants’ willingness to engage in corrupt practices in Costa Rica. Conversely, Agerberg (2022) offered a more optimistic perspective from Mexico, showing that messages emphasizing widespread public rejection of corruption significantly improved interpersonal trust and reduced willingness to participate in bribery.

Building on this evidence and recognizing the theoretical importance of priors in shaping how individuals interpret information, our study adopts a relatively unobtrusive “debiasing” approach. By providing an external benchmark, we aimed to shift individuals’ attention away from entrenched biases and toward a more objective assessment of corruption. The “debiasing” and “sequential” treatments were effective in reducing perceptions of corruption and inclinations toward tax evasion. Among the treatments, the “sequential” approach proved most effective, reducing corruption perceptions by 13% relative to the baseline and significantly lowering the likelihood of tax evasion. This finding was further validated in our field experiment, where the “sequential” treatment increased declared taxable income, particularly among high-risk taxpayers. The “reform” treatment, which focused solely on highlighting government actions, failed to produce significant effects and even backfired in some cases.

Our findings carry significant implications for policymakers. In low-trust contexts,

exaggerated beliefs about corruption can lead anticorruption messages to inadvertently reinforce negative perceptions, while skepticism about government intentions may result in outright rejection of such efforts. These challenges are particularly acute in developing countries, where weak institutions and widespread tax evasion undermine governance. For instance, an estimated 25% of total sales in African and Latin American countries go unreported for tax purposes, compared to just 7% in OECD countries (Kouame and Goyette, 2018). Addressing distorted beliefs through theoretically grounded interventions can not only help to rebuild trust in tax administrations, but also foster voluntary compliance, thus enhancing revenue collection.

Our study demonstrates that beliefs about corruption are both critical and malleable. Effective communication in anticorruption campaigns requires a two-step approach: first, altering perceptions of corruption, and second, emphasizing government achievements. By carefully sequencing these messages, governments can foster more positive views of institutional competence and reduce tendencies toward tax evasion or corrupt behaviors. These insights highlight the importance of integrating behavioral considerations into the design of anticorruption strategies, offering a more optimistic perspective on the potential for communication campaigns to drive meaningful change.

References

- Acemoglu, D. and Verdier, T. (2000). The choice between market failures and corruption. *American Economic Review*, 91(1):194–211.
- Adida, C., Gottlieb, J., Kramon, E., and McClendon, G. (2017). Overcoming or reinforcing coethnic preferences? an experiment on information and ethnic voting. *Quarterly Journal of Political Science*, 12(4):437–477.
- Agerberg, M. (2022). Messaging about corruption: The power of social norms. *Governance*, 35(3):929–950.
- Ajzenman, N. (2021). The power of example: Corruption spurs corruption. *American Economic Journal: Applied Economics*, 13(2):230–257.
- Alm, J., Martinez-Vazquez, J., and McClellan, C. (2016). Corruption and firm tax evasion. *Journal of Economic Behavior & Organization*, 124:146–163.
- Antinyan, A. and Asatryan, Z. (2024). Nudging for tax compliance: a meta-analysis. *The Economic Journal*, page ueae088.
- Barabas, J. (2004). How deliberation affects policy opinions. *American Political Science Review*, 98(4):687–701.
- Barrera, O., Guriev, S., Henry, E., and Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics*, 182:104123.
- Baum, A., Gupta, S., Kimani, E., and Tapsoba, S. (2024). Citizens Perceptions of Tax Authorities and Tax Efficiency. *IMF Working Papers*, 2024(234):1.
- Beesley, C. and Hawkins, D. (2022). Corruption, institutional trust and political engagement in Peru. *World Development*, 151:105743.
- Besley, T. and Mueller, H. (2021). Fiscal capacity and state fragility. In *Macroeconomic Policy in Fragile States*. Oxford University Press.

- Bottan, N. L. and Perez-Truglia, R. (2020). Betting on the house: Subjective expectations and market choices. Technical report, National Bureau of Economic Research.
- Boudreau, C. and MacKenzie, S. A. (2014). Informing the electorate? how party cues and policy information affect public opinion about initiatives. *American Journal of Political Science*, 58(1):48–62.
- Broockman, D. E. and Kalla, J. L. (2023). When and why are campaigns’ persuasive effects small? Evidence from the 2020 US presidential election. *American Journal of Political Science*, 67(4):833–849.
- Bullock, J. G. (2009). Partisan bias and the bayesian ideal in the study of public opinion. *The Journal of Politics*, 71(3):1109–1124.
- Bullock, J. G. (2011). Elite influence on public opinion in an informed electorate. *American Political Science Review*, 105(3):496–515.
- Bullock, J. G., Gerber, A. S., Hill, S. J., and Huber, G. A. (2015). Partisan bias in factual beliefs about politics. *Quarterly Journal of Political Science*, 10(4):519–578.
- Bénabou, R. and Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–64.
- Cheeseman, N. and Peiffer, C. (2022). The curse of good intentions: Why anticorruption messaging can encourage bribery. *American Political Science Review*, 116(3):1081–1095.
- Cheeseman, N. and Peiffer, C. (2023). Why efforts to fight corruption can undermine the social contract: Lessons from a survey experiment in Nigeria. *Governance*, 36(4):1045–1061.
- Chong, A., De La O, A. L., Karlan, D., and 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.

- Christensen, L. (2023). Optimal persuasion under confirmation bias: Theory and evidence from a registered report. *Journal of Experimental Political Science*, 10(1):4–20.
- Cieřlik, A. and Goczek, Ł. (2018). Control of corruption, international investment, and economic growth—evidence from panel data. *World Development*, 103:323–335.
- Cilliers, J., Elashmawy, N. A. M. A., and McKenzie, D. J. (2023). Using post-double selection lasso in field experiments. Policy research working paper, World Bank Group, Washington, D.C.
- Corbacho, A., Gingerich, D. W., Oliveros, V., and 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.
- Cotter, R. G., Lodge, M., and Vidigal, R. (2020a). The boundary conditions of motivated reasoning. In et al., E. S., editor, *The Oxford Handbook of Electoral Persuasion*, pages 66–87. Oxford University Press.
- Cotter, R. G., Lodge, M., and Vidigal, R. (2020b). When, how, and why persuasion fails: A motivated reasoning account. In et al., E. S., editor, *The Oxford Handbook of Electoral Persuasion*, pages 51–65. Oxford University Press, Oxford.
- Cullen, Z. and Perez-Truglia, R. (2022). How much does your boss make? The effects of salary comparisons. *Journal of Political Economy*, 130(3):766–822.
- Daniele, G., Aassve, A., and Le Moglie, M. (2023). Never forget the first time: The persistent effects of corruption and the rise of populism in Italy. *The Journal of Politics*, 85(2):468–483.
- DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2):315–72.
- Dunning, T. (2012). *Natural experiments in the social sciences: A design-based approach*. Cambridge University Press.

- Enríquez, J. R., Larreguy, H., Marshall, J., and Simpser, A. (2024). Accountability under polarization.
- Epley, N. and Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, 30(3):133–40.
- Erlich, A. and Gans-Morse, J. (2025). Can norm-based information campaigns reduce corruption? *American Journal of Political Science*, pages 1–20.
- Fisman, R. and Svensson, J. (2007). Are corruption and taxation really harmful to growth? firm level evidence. *Journal of Development Economics*, 83(1):63–75.
- Flynn, D. J., Nyhan, B., and Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38:127–150.
- Guess, A. and Coppock, A. (2020). Does counter-attitudinal information cause backlash? results from three large survey experiments. *British Journal of Political Science*, 50(4):1497–1515.
- Gulino, G. and Masera, F. (2023). Contagious dishonesty: Corruption scandals and supermarket theft. *American Economic Journal: Applied Economics*, 15(4):218–251.
- Haaland, I., Roth, C., and Wohlfart, J. (2023). Designing Information Provision Experiments. *Journal of Economic Literature*, 61(1):3–40.
- Hewitt, L., Broockman, D., Coppock, A., Tappin, B. M., Slezak, J., Coffman, V., Lubin, N., and Hamidian, M. (2024). How experiments help campaigns persuade voters: Evidence from a large archive of campaigns’ own experiments. *American Political Science Review*, pages 1–19.
- IDB (2015). Tax Administration Institutional and Operational Strengthening. Technical report, Inter-American Development Bank.
- Iyengar, S. and Kinder, D. R. (2010). *News that matters: Television and American opinion*. University of Chicago Press.

- Jann, B., Jerke, J., and Krumpal, I. (2012). Asking sensitive questions using the cross-wise model: An experimental survey measuring plagiarism. *Public Opinion Quarterly*, 76(1):32–49.
- Kouame, W. A. K. and Goyette, J. (2018). Tax evasion in africa and latin america: the role of distortionary infrastructures and policies. *World Bank Policy Research Working Paper*, (8522).
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3):480.
- Levi, M. (1988). *Of rule and revenue*. University of California Press.
- Levi, M., Sacks, A., and Tyler, T. (2009). Conceptualizing legitimacy, measuring legitimating beliefs. *American Behavioral Scientist*, 53(3):354–375.
- Little, A. T., Schnakenberg, K. E., and Turner, I. R. (2022). Motivated reasoning and democratic accountability. *American Political Science Review*, 116(2):751–767.
- Luttmer, E. F. P. and Singhal, M. (2014). Tax morale. *Journal of Economic Perspectives*, 28(4):149–68.
- Martin, L. E. (2023). *Strategic Taxation: Fiscal Capacity and Accountability in African States*. Oxford University Press.
- Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, 110(3):681–712.
- Nicholson, S. P. (2012). Polarizing cues. *American Journal of Political Science*, 56(1):52–66.
- Nyhan, B. and Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2):303–330.
- Okunogbe, O. and Tourek, G. (2024). How can lower-income countries collect more taxes? The role of technology, tax agents, and politics. *Journal of Economic Perspectives*, 38(1):81–106.

- Olken, B. A. (2006). Corruption and the costs of redistribution: Micro evidence from Indonesia. *Journal of Public Economics*, 90(4):853–870.
- Peiffer, C. (2020). Message received? experimental findings on how messages about corruption shape perceptions. *British Journal of Political Science*, 50(3):1207–1215.
- Peiffer, C. and Alvarez, L. (2016). Who will be the “principled-principals”? Perceptions of corruption and willingness to engage in anticorruption activism. *Governance*, 29(3):351–369.
- Pomeranz, D. and Vila-Belda, J. (2019). Taking state-capacity research to the field: Insights from collaborations with tax authorities. *Annual Review of Economics*, 11(Volume 11, 2019):755–781.
- Rabin, M. and Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics*, 114(1):37–82.
- Roth, C. and Wohlfart, J. (2020). How do expectations about the macroeconomy affect personal expectations and behavior? *Review of Economics and Statistics*, 102(4):731–748.
- Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*, 57(4):904–54.
- Slothuus, R. and De Vreese, C. H. (2010). Political parties, motivated reasoning, and issue framing effects. *The Journal of Politics*, 72(3):630–645.
- Taber, C. S., Cann, D., and Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior*, 31:137–155.
- Taber, C. S. and Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3):755–769.
- Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157):1124–1131.

- Weigel, J. L. and Kabue Ngindu, E. (2023). The taxman cometh: Pathways out of a low-capacity trap in the Democratic Republic of the Congo. *Economica*, 90(360):1362–1396.
- Weitz-Shapiro, R. and Winters, M. S. (2017). Can citizens discern? Information credibility, political sophistication, and the punishment of corruption in Brazil. *The Journal of Politics*, 79(1):60–74.
- Yair, O., Sulitzeanu-Kenan, R., and Dotan, Y. (2020). Can institutions make voters care about corruption? *The Journal of Politics*, 82(4):1430–1442.
- Yamou, T., Thomas, A. H., and Cai, K. (2024). Citizens Perceptions of Tax Authorities and Tax Efficiency. *IMF Working Papers*, 2024(234):1.
- Zaller, J. (1992). The nature and origins of mass opinion. *Cambridge University*.