Voice and Punishment

A Global Survey Experiment on Tax Morale

Fredrik M. Sjoberg
Jonathan Mellon
Tiago Peixoto
Johannes Hemker
Lily L. Tsai

WORLD BANK GROUP

Governance Global Practice
May 2019

Abstract

An online survey experiment spanning 50 countries finds sizable improvements in tax morale when (a) the salience of anti-corruption efforts is increased and (b) citizens are allowed to voice their expenditure preferences to the government. These results hold very broadly across a uniquely large and diverse sample of respondents from all continents.

The findings are consistent with theories emphasizing the role of democratic accountability, as well as of perceptions of legitimacy and "retributive justice," in generating voluntary tax compliance. Implications and avenues for further research are discussed.

Voice and Punishment:
A Global Survey Experiment on Tax Morale

Fredrik M. Sjoberg
Jonathan Mellon
Tiago Peixoto
Johannes Hemker
Lily L. Tsai

Introduction

Taxation is a key, and perhaps the most foundational, element of the relationship between citizens and governments. For citizens, the institution of taxation of course has important effects on disposable income and publicly available goods and services. For policy makers and governments, raising revenues to fund government activity is a necessary condition for assuring institutional survival and maintaining public order. The question of why and under what conditions citizens comply with tax demands made on them by governments is therefore central to our understanding both of tax policy, and of political economy and governance more generally.

Over the last decades, scholars and practitioners have increasingly acknowledged that tax compliance can only partially be understood as a narrowly "rational" act in the sense that taxpayers only comply because the expected costs of avoiding or evading taxes exceed the costs of taxation. In this context, the term *tax morale* has often come to denote – basically as a residual category – motivations for complying with taxes that go beyond the expected costs of detection and punishment in models of "rational crime". Research has unearthed considerable variation in tax morale between individuals, regions and countries, and has generated useful theory and tools for understanding it.

The key questions about tax morale for policy makers and governments relate to their ability to affect it. Higher tax morale is an asset for governments on many levels: lower tax collection costs and higher revenues have direct financial value (conditional on policy). Tax morale can also benefit economic policy indirectly. For example, if higher nominal tax rates increase the marginal distortions and welfare costs of taxation and higher compliance allows for lower nominal rates, tax morale can reduce the welfare burden of taxation. Higher tax morale may also allow developing country governments to rely less on tariffs, which are often thought to be especially distortionary and detrimental to growth.

These considerations lead to the empirical question of what kinds of motivations generate tax morale, and how they can best be promoted. A growing literature has sought to address this question, and found some experimental evidence linking different types of interventions to tax morale in a variety of settings. However, since much of this literature has focused on a small set of developed countries, it is very unclear how broadly applicable these findings are, and how universally the suggested mechanisms apply.

In this paper, we use a new type of survey experiment implemented with a uniquely broad sample of 65,000 respondents from 50 countries to investigate this question. Specifically, we estimate the effects of two interventions on tax morale: a "bottom up participation" intervention asking respondents to state their expenditure preferences, and a "top down accountability" intervention giving respondents information about anti-corruption agencies and the punishments they help inflict on corrupt officials. The design was pre-registered with EGAP prior to data collection.

Overall, we find large and statistically significant positive effects of both interventions on two measures of tax morale. These results are basically unaffected by a series of demanding robustness checks. We are also able to demonstrate that subjects actually absorb the

information given to them in the interventions, and that women respond less strongly to them than men.

The contribution of the paper is twofold. First, the unusually large sample with respondents from all continents allows us to assess the causal effects across a wide range of contexts, and we find that effects are quite homogeneous. This suggests previous research on the effectiveness of participation interventions in increasing tax morale might be widely applicable, even when the interventions are relatively shallow in the sense that they are short-lasting and cheap to implement. Second, the new intervention based on increasing the salience of anti-corruption efforts proves to be highly effective at increasing tax morale. In our view, this constitutes novel and rigorous evidence that "retributive justice", i.e. the sense that governments systematically punish abuse of public money and corruption is an important driver of tax morale.

The remainder of the paper is structured as follows: The theory section reviews previous research motivating our study, followed by a section which introduces our research design and survey instruments. In the next and main section, we review results and perform robustness checks. The final section concludes with a discussion of implications and future research.

Theory

Rational tax compliance

The baseline framework for thinking about tax compliance is from Allingham and Sandmo (1972), who model tax reporting decisions from the perspective of utility maximizing agents considering a trade-off between the costs of paying taxes and the costs of being caught cheating. This model gives rise to a number of hypotheses, notably about the ceteris paribus effect of tax rates, enforcement efforts and sanctions on observed tax compliance, which have been the subject of considerable empirical scrutiny over decades (Slemrod and Yitzhaki 2002).

Recent studies in this vein include Castro and Scartascini (2015), who present experimental evidence that people respond to deterrent messaging, but with significant variation depending on the channels of communication. In their study, deterrence messages conveyed by inspectors were the most effective, followed by those sent via email, while those sent via letter were least effective. Moreover, there is survey evidence suggesting a link between deterrence and tax morale (D'Arcy 2009; Ali, Fjeldstad, and Sjursen 2014; Yesegat and Fjeldstad 2016; Fjeldstad, Schulz-Herzenberg, and Hoem Sjursen 2012; Gobena and Van Dijke 2016).

On the other hand, Del Carpio (2013) conducted a field experiment in Peru and found that taxpayers who received information on the level of enforcement were not more likely to comply than taxpayers who were simply reminded of their duty to comply.

But even if one grants that some comparative statics from the Allingham and Sandmo model have a solid empirical basis, some important descriptive facts about tax compliance remain unaccounted for by such a model. First, as early work by Alm, McClelland, and

Schulze (1992) shows, observed tax compliance is much higher in most contexts than one would expect based on plausible values of detection probabilities, sanction costs, and individual risk aversion. Second, in a more extreme example, there appears to be compliance even with taxes that are not enforced at all, casting doubt on the role of detection and deterrence in tax compliance. As Dwenger et al. (2016) find based on a local church tax in Germany, individuals pay taxes even when there are no enforcement efforts at all. And, consistent with the notion that individuals are in fact aware of the lack of enforcement, letters informing taxpayers that there is no enforcement do not appear to influence taxpayer behavior. Third, as Luttmer and Singhal (2014) point out, governments across the world appear to be investing in increasing individuals' intrinsic motivation to pay taxes, and cultivating and reinforcing norms of tax compliance, for example by publishing lists of tax debtors or recognizing "distinguished taxpayers".

For these reasons, social scientific work has increasingly investigated tax morale as a factor in tax compliance. Below, we focus on the aspects of reciprocity, procedural justice and institutions that our study speaks most directly to.

Tax morale, reciprocity and fiscal exchange

One straightforward reason why individuals might voluntarily comply with taxes is because they view taxes as part of a social contract where they help fund the public purse and the state provides services in return. This logic of reciprocity is at the core of the "fiscal exchange" or "social contract" hypothesis, which stipulates that the link between taxes and government spending helps explain compliance.

There is some empirical evidence in favor of this hypothesis. Flores-Macías (2016) uses experimental evidence from Mexico to show that respondents who were reminded that their tax money is spent on security (the most salient public good in the country) show the highest increase in tax compliance, compared to other messages. Ortega, Ronconi, and Sanguinetti (2016) also provide experimental evidence from Latin America showing that taxpayers who were given vignettes about high government performance in the provision of public goods like school supplies were more tax compliant. Ali, Fjeldstad, and Sjursen (2014) find that in Tanzania, Uganda, Kenya, and South Africa, the level of public goods provision has a significant, positive relationship with tax compliance.

Other studies present evidence that is less aligned with the fiscal exchange hypothesis. Yesegat and Fjeldstad (2016) find no evidence for a logic of fiscal exchange among business-owners in Ethiopia, although measurement is poor and the sample is business-owners, who might behave differently from private individuals. The experiment in Castro and Scartascini (2015) finds null effects on compliance when respondents are reminded about public goods provision, although the authors attribute this to a weak treatment and the difficulty of changing perceptions of goods provision with just one message.

Tax morale, institutions and political legitimacy

A related but distinct argument to the fiscal exchange hypothesis is that individuals comply with taxes because they trust the government and think that the process by which the tax was decided on is fair and legitimate.

Descriptive cross-country work based on individual-level surveys such as the World Values Survey shows that trust in government, support for democracy, and measures of the quality of institutions are positively related to survey-based measures of tax morale (Alm and Torgler 2006). Survey-based research focusing on the developing world by Daude, Gutiérrez, and Melguizo (2012) finds some support for a positive relationship between tax compliance and trust in government as well. On the other hand, the correlation between tolerance of tax evasion and the Polity IV index of democracy does not appear as clear-cut in developing countries as some studies would have it (Xiaobo et al. 2018). With this caveat, and although it would be a stretch to interpret these studies causally, cross-national survey studies have provided an important basis and motivation for more (quasi-) experimental studies into the determinants of tax morale.

In observational studies, Pommerehne and Weck-Hannemann (1996) and Torgler (2005) show that Swiss cantons with more direct-democratic institutions also exhibit higher tax compliance. A similar implication is generated in the study by Alm and Torgler (2006) which finds that Switzerland and the U.S. have both strong direct democratic institutions and among the highest levels of tax compliance internationally. Much of the experimental work in this area has focused on laboratory settings, where many aspects of the environment are controlled by the researcher and the context and information environment can be manipulated more precisely. In one early landmark laboratory study Alm, Jackson, and McKee (1992) found that making decisions by voting increased tax compliance significantly in the lab. In their study of Swiss college students, Feld and Tyran (2002) argue that only notions of procedural legitimacy, which is generated by having laboratory subjects vote on a fine, can explain their finding that subjects comply more with a fine when it is voted on than when it is imposed. Importantly, while they do find a role for reciprocity (compliance with the fine is higher the more subjects voted for it) in explaining tax compliance, their argument goes beyond pure reciprocity to focus on *procedural fairness*, or the belief that the method by which the decision was reached was fair.

More recently, Lamberton, De Neve, and Michael (2014) found evidence, also in a lab setting using American respondents, showing that simply allowing respondents to state their expenditure preferences reduced their use of a questionable tax loophole, thus raising compliance. They interpret this finding as reflecting a re-coupling of taxes with the services that they fund along the lines of the fiscal exchange hypothesis.

Finally, in a study based on Austrian laboratory subjects, Casal et al. (2016) find that a different type of "voice" also matters: subjects randomly assigned to a condition where they can make item-by-item decisions on which taxes to pay have higher total tax compliance than subjects randomly assigned to paying all taxes together.

However, while these types of studies credibly isolate causal effects, it is very difficult to get a sense of how much they hinge on the specific setting. First, while laboratory experiments do give researchers a large amount of control, they also tend to involve very specific types

of respondents in artificial situations, and it is at least unclear how well they generalize to real-world behavior.

Second, and perhaps even more importantly, the vast majority of experimental studies of tax morale are implemented in rich Western democracies. Given that tax compliance is deeply interwoven with aspects of national tax and budget policy, political institutions and political culture, this raises important questions of the conditions, if any, under which we might expect to see the positive effects described above replicated elsewhere.

The importance of such issues of generalizability becomes especially visible when researchers use interventions tested in laboratory settings, and apply them to very specific real-world settings. For example, in a recent pioneering study, Kettle et al. (2017) randomly assign Guatemalan taxpayers to six different messages during their tax declaration, and fail to find any significant effects. Of course, in a sense, this is only surprising if one was previously led to believe that Guatemalan VAT tax payers should respond similarly to interventions as Swiss college students in the laboratory. The null results are just as consistent with laboratory experiments being generally wrong as they are with Guatemalan taxpayers being an exception from the rule.

In our view, this is an illustration of why there is value to a more large-scale and cross-country experimental study of tax morale. As described below, we implement our study on a uniquely broad set of respondents from countries spanning all continents, and exhibiting a wide variety of political institutions and tax systems. We believe that seeing how one and the same experiment replicates across a vast set of contexts is informative both for researchers interested in explanations that are as general as possible and for policy makers interested in finding solutions that are as specific as necessary.

Retributive justice and tax morale

Finally, beyond specifically political institutions, specific arrangements related to the proper administration of government funds may also matter for citizens' willingness to contribute to the tax system. In particular, as Tsai (2017) and Tsai et al. (2019) argue, high-level institutions that punish malfeasance in the public sector can help signal to ordinary citizens that their government cares about the public interest, and that corruption is not tolerated. As she argues, such institutions, besides deterring malfeasance ex ante, also help uphold the fundamental values of the political community, and allow elites to show themselves to be moral actors and leaders.

In this vein, Xiaobo et al. (2019) show in an empirical analysis of property taxes in China that the effectiveness of citizen input in generating citizens' willingness to comply with taxation is dependent upon the existence of top-down sanctioning institutions. Below, we use a new prime about the existence of such institutions of "retributive justice"" as a treatment to investigate whether these affect tax morale.

Research Design

In order to contribute to a better understanding of tax morale, we conduct a global online survey experiment of tax morale attitudes. At the empirical core of the study, we use an

experiment to estimate the effects of two interventions on tax morale, relative to a control group:

First, we look at the effects of a bottom-up participation treatment, where we ask respondents to state their spending preferences. As discussed above, treatments eliciting spending preferences have been found to increase tax morale in some experimental literature, though much of the research here has been confined to specific political and cultural contexts. The breadth of our global survey experiment allows us to get a better sense of the degree to which this type of intervention actually affects tax morale.

Second, in order to test the predictions of a theory of retributive justice, we look at whether a "top-down accountability" treatment, which increases the salience of anti-corruption agencies and the punishments they inflict on corrupt officials, can improve tax morale.

To implement this test, we randomly vary a portion of the survey between the three conditions as laid out in Table 1.

Table 1: Treatments

Control group	Top-down accountability intervention	Bottom-up participation intervention
There are many popular search engines in the world, with different designs and functions. Search engines are used every day by over 1 billion people worldwide.	When government money is misused, it is very important to find and punish those responsible. Your government has a national agency, the [AGENCY NAME], that helps to punish the misuse of government funds.	You have been selected to be part of the Online Citizen Assembly: a national conversation about how the government of [COUNTRY NAME] should spend money. The results of the Online Citizen Assembly will be presented to the government.
Many people say they are annoyed by all the advertising on search engines. How much does this apply to you?	The [AGENCY NAME] has investigated many cases of government corruption. Many people who misused government funds have been punished. Do you think it is good to have an agency that investigates government corruption?	What should your government spend more money on? The results of the Online Citizen Assembly will be presented to the government.
Not at all A bit Quite a bit Very much	Yes No	Defense and Police Education Transportation Welfare Health Environment

To measure tax morale in the survey, we use two outcome questions.

1) The tax morale question from the General Social Survey (GSS):

If a taxpayer does not report all of their income in order to pay less income taxes do you feel it is: (Not wrong / A bit wrong / Wrong / Seriously wrong)

and,

2) A tax fine attitude question:

If a taxpayer does not report all of their income in order to pay less income taxes, what percentage of their income should they pay as a penalty? (None (0%) / 1-10% / 11-20% / More than 20%)

Although we think of both of these questions as tapping tax morale, note that there are important differences here between the first question, which is framed as a moral question, and the second question, which is more specific and boils down to an actual policy parameter. Also note that, while many might agree that increasing tax morale is desirable in most contexts, it is not necessarily clear that increasing punitive preferences in society is desirable in and of itself.

We also ask questions about basic covariates (age, gender, labor market status), as well as "manipulation check" and "mechanism" questions that allow us to look at whether respondents absorb the information given to them and investigate potential mechanisms. The full survey can be found in the appendix.

This survey was served to internet users in the designated countries using an online survey platform (Riwi) which uses frequently-occurring typos in websites (e.g. googel.com) to generate impressions. This means that respondents do not expect to take the survey, and can choose to cancel at any point in time. Platforms such as Riwi have huge advantages compared to traditional survey methods in terms of cost, speed and geographical coverage, but are more prone to issues such as attrition since respondents were not expecting to take a survey in the first place.

The full list of countries in which we ran the experiments, by region, is reproduced below.

Box: Countries in sample

Africa: Algeria, Morocco, Tunisia, South Africa, Kenya, Nigeria, Uganda, Angola

Middle East: Arab Republic of Egypt, Jordan, Saudi Arabia

Europe: Spain, Czech Republic, Netherlands, Belgium, Austria, Germany, Greece, Hungary, Italy, Kazakhstan, Poland, Portugal, Romania, Russian Federation, Slovak Republic, Sweden, Turkey

Asia: Thailand, Vietnam, Bangladesh, China, Taiwan, China, Australia, India, Philippines, Indonesia, Japan, Republic of Korea, Malaysia, Pakistan

Americas: United States, Canada, Brazil, Argentina, Chile, Colombia, Mexico, Peru, Venezuela RB

The questionnaire was translated into the local language by professional translators, and then checked by researchers who were native speakers to ensure quality. Respondents in multilingual countries were able to pick a language.

Prior to gathering the data, we pre-registered our hypotheses and methodology in a pre-analysis plan with EGAP. In short, we decided to look at the sum of the tax morale questions as a key variable (but also analyze them separately afterwards), expected both treatments

to improve tax morale relative to the control group, laid out our robustness checks, and recorded some specific hypotheses about heterogeneous treatment effects and potential mechanisms in the pre-analysis plan. That analysis is implemented below.

Results

Descriptive Statistics

The study was conducted during three weeks in the summer of 2017. A total of 151,096 subjects answered Q1 (after the age & gender selector) and 65,471 answered Q9, a completion rate of 43.3%.

Descriptive statistics from the subjects that completed the study are shown in Table 2.

Table 2: Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Max
Age	65,436	33.53	14.14	16	65
Female	65,435	0.30	0.46	0	1
Business Owner	65,432	0.07	0.25	0	1
Smartphone	65,436	0.39	0.49	0	1
Tax fine outcome	65,430	2.66	1.05	1	4
Tax morale outcome	65,429	5.19	1.86	2	8
Tax index	65,433	0.49	0.50	0	1
Top-down manipulation check	65,434	0.53	0.50	0	1
Bottom-up manipulation check	65,435	0.35	0.48	0	1
Politicians care ==1	65,435	0.71	0.45	0	1

The distributions of responses to the two outcome questions are shown in Figure 1.
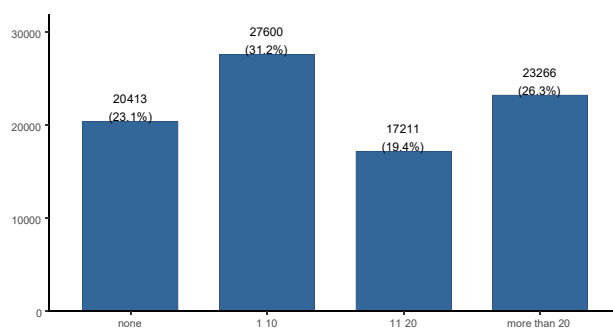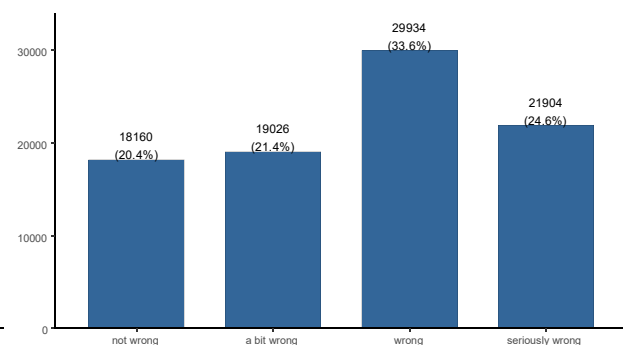

Figure 1: Distribution of responses to the two main outcome questions, tax morale and tax fine

The two outcome variables are correlated at 0.495. Table 3 displays the relationship between them. Of course, the fact that the instruments are only moderately correlated can be attributed both to measurement error and/or to actual differences in what they measure. However, note that all of our key results hold for both outcome measures.

Table 3: Relationship between the tax fine and tax morale outcome variables. Cell percentages shown (i.e. the numbers across the whole table sum to 1). There is a moderately strong positive relationship between the two dependent variables.

Q4 (Penalty) / Q5 (Evasion is...)	not wrong	a bit wrong	wrong	seriously wrong
None	0.116	0.038	0.035	0.020
1-10%	0.038	0.119	0.127	0.042
11-20%	0.012	0.035	0.107	0.042
> 20%	0.024	0.02	0.076	0.149

Treatment effects without covariate adjustment

As we had specified in the pre-analysis plan, our first cut at estimating treatment affects is a raw comparison of means across conditions, pooled across all countries, without any covariate adjustment. As Figure 2 shows, the treatments appear to have had a large and significant effect on the main outcome variable, the sum of the two outcome questions. On a scale of 1 to 8, tax morale appears highest in the "top-down accountability" condition at 5.30, followed by the bottom up participation condition at 5.23, and is lowest in the control condition at 5.07.
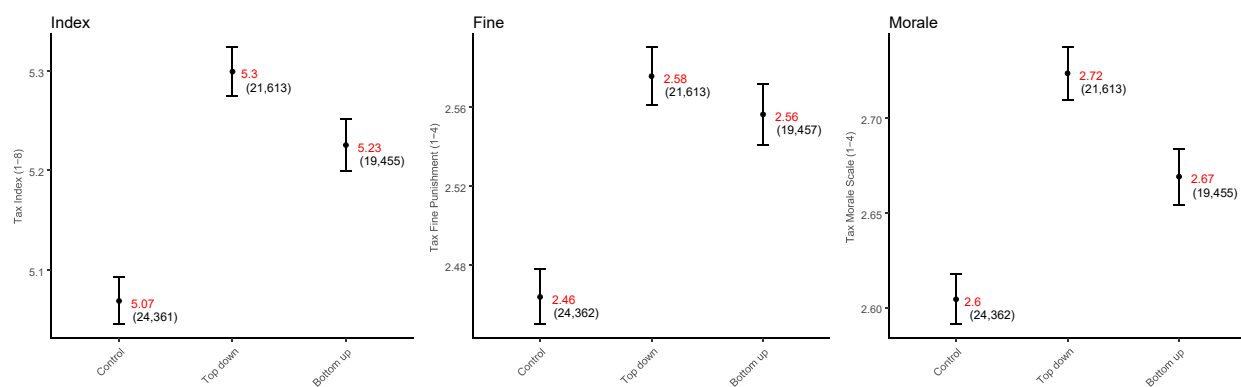
Figure 2: Simple comparison of means, Combined index, tax fine and, Tax Morale (GSS) (95% CIs)

This effect is sizable in magnitude, though not extremely large relative to the underlying variation in tax morale: in the top-down accountability condition, tax morale is 0.124 standard deviations higher than in the control group. Looking at the two component parts of the tax index, the rank order between conditions remains the same, though the difference

between the two treatment conditions is a bit smaller on the tax fine attitude measure than on the GSS tax morale question.

It is difficult to directly compare these results to those in previous experimental studies on tax compliance, as many of those use binary behavioral indicators. However, these measures have been used in a number of observational studies. The effect sizes are somewhat larger than that observed for a 1 point increase in direct democratic rights on a 1-6 scale or a 1 point increase in trust in the legal system on a 1-5 scale (Torgler, 2005). In other words, these effect sizes are comparable to those effects seen for meaningful increases in institutional trust and involvement in observational studies.

Cross-country variation

While we did not have ex ante hypotheses or strong priors about cross-country patterns of effects, it is also useful to look at estimates at the country level to make variation, as well as the overarching pattern, transparent and visible. We first show histograms (Figure 3) of country-level estimated effects for the two treatments. This shows that the while there is quite a bit of cross-country variation in estimated ATEs, the mean effect estimate is not driven by extreme outlier countries. This is the case in particular with regard to the top-down treatment.
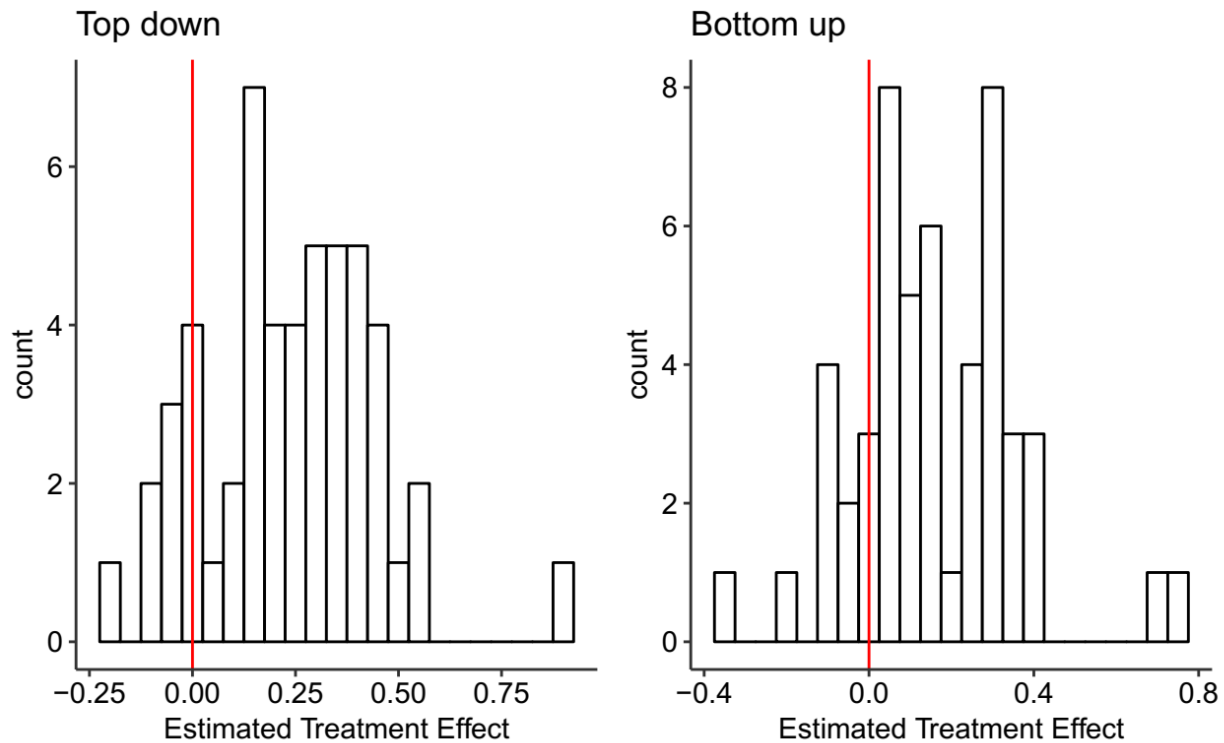

Figure 3: Distribution of country-level ATEs, for Top Down and Bottom-Up-Treatment for Tax Index Outcome

Next, we correlate the estimated effects from the two treatments with each other at the country level (Figure 4). By and large, respondents from countries with large effects of one

treatment are also estimated to exhibit large effects of the other. Finally, for completeness, we plot means across all conditions for all countries in Figure 7 in the appendix.
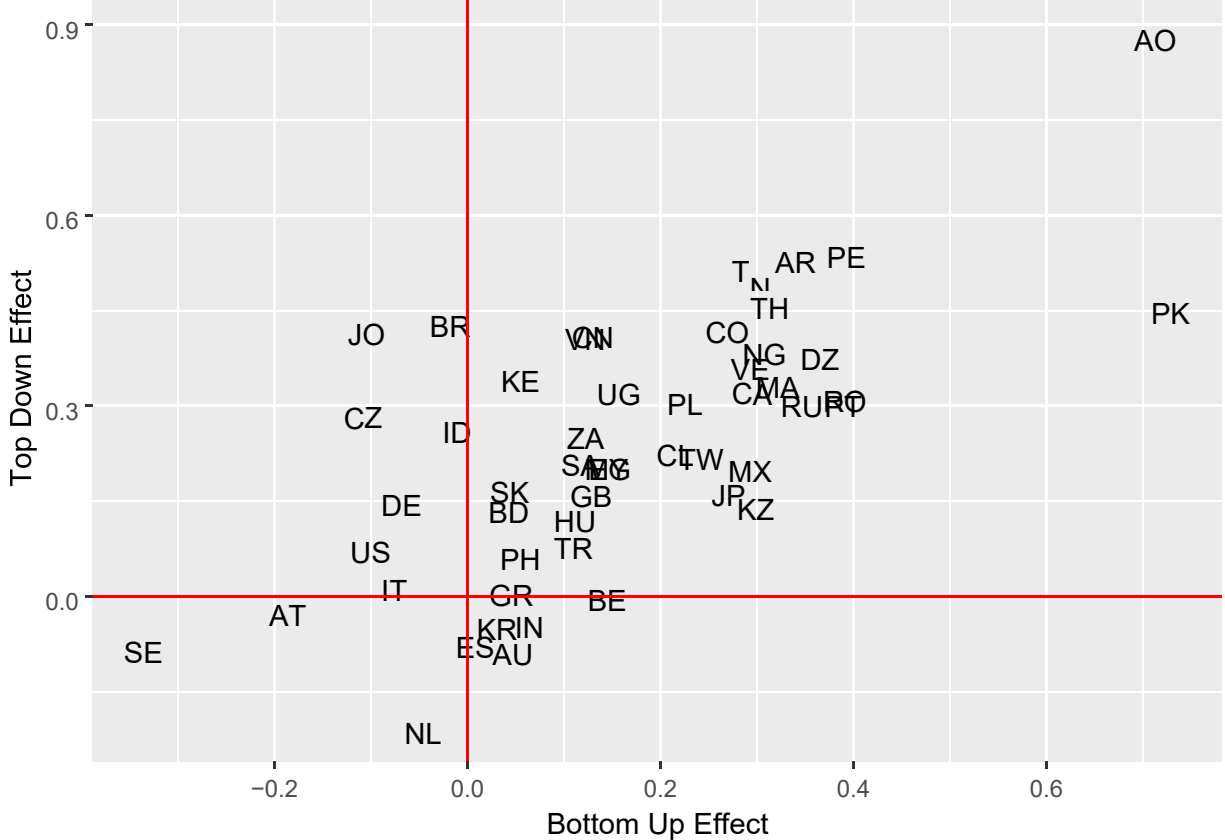
Figure 4: Bottom-up and Top-down Treatment Effects, by country

Estimating treatment effects with covariate adjustment

Next, we run linear models with and without covariate adjustments: as per the pre-analysis plan, we use simple OLS regressions of the outcome variables (the index or the two constituent parts) on the treatment indicators. In models with covariate adjustment, we include the pre-treatment covariates: age buckets, gender, and employment status.

As Table 4 shows, covariate adjustments do not change the estimated treatment effects appreciably.

Table 4: Average Treatment Effects with and without Covariate Adjustment (OLS)

	Tax index raw	Tax index controls	Tax morale raw	Tax morale controls	Tax fine raw	Tax fine controls
Top down/Control	0.231***	0.231***	0.119***	0.120***	0.111***	0.111***
	(0.017)	(0.017)	(0.010)	(0.010)	(0.010)	(0.010)
Bottom up/Control	0.157***	0.164***	0.065***	0.069***	0.092***	0.095***
	(0.018)	(0.018)	(0.010)	(0.010)	(0.011)	(0.010)

N	65429	65426	65430	65427	65432	65429
p	0.0	0.0	0.0	0.0	0.0	0.0
Log-likelihood	-133337.7	-132601.4	-96130.6	-95363.8	-98885.6	-98487.5
AIC	266683.4	265230.7	192269.3	190755.5	197779.2	197003.1

Notes: each column shows an OLS model predicting a dependent variable with or without controls for age, gender and employment status. Tax index is the sum of the tax morale and tax fine variables.

Manipulation checks

We run models to assess whether the treatments affect responses to the manipulation check questions. The questions were phrased as follows:

> *What type of corruption does the [comptroller name here] investigate? (Government / Private sector / Not sure)*

And,

> *After taking this survey, do you feel like you have opportunities to tell the government how to spend money? (Yes / No)*

Since these questions were designed to make sure that the treatments would affect them, failure to reject the null hypothesis in these models would lead us to question the effectiveness of the treatment. As Table 5 shows, the treatments are effective in moving respondents' answers to the manipulation check questions. However, there appears to be some spillover between treatments: the bottom-up treatment also appears to move the top-down manipulation check and vice versa, though only to about half the degree that the "correct" treatment affects the manipulation checks. We also implement a robustness check in Table 6 based on the manipulation check results.

Table 5: Manipulation check models

	Top-down manip check	Bottom-up manip check
(Intercept)	0.462***	0.480***
	(0.003)	(0.003)
Top down/Control	0.064***	0.055***
	(0.005)	(0.005)
Bottom up/Control	0.031***	0.091***
	(0.005)	(0.005)
N	65433	65434
p	0.0	0.0
Log-likelihood	-47390.0	-47224.7

AIC	94788.1	94457.5

Notes: Each column shows an OLS model predicting the manipulation check variables relating to the top down treatment and the bottom up treatment. The relevant treatment has a large effect on the correct manipulation check in each case.

Robustness checks

Next, we implement robustness checks that were outlined in the pre-analysis plan to assess the robustness of our findings. For each of the three outcomes, we first show the baseline model with covariate adjustments in the first column, followed by

- Balance: A model estimated only on the subset of countries where we cannot reject balance for any covariate at the .05 level
- Attrition: A model estimated only on the subset of countries where treatment status does not predict completing the survey
- Attrition (cond): A model estimated only on the subset of countries where, jointly with covariates, treatment status does not predict completing the survey
- Bottom manip: A model estimated only on the subset of countries where we can reject the hypothesis that the bottom-down treatment had no effect on the corresponding manipulation check
- Top manip: A model estimated only on the subset of countries where we can reject the hypothesis that the top-down treatment had no effect on the corresponding manipulation check

Table 6: Main Model and Robustness Checks: Tax Index (OLS)

	Main	Balance	Attrition	Attrition (cond)	Bottom manip	Top manip
Top down/Control	0.231***	0.235***	0.273***	0.273***	0.265***	0.223***
	(0.017)	(0.018)	(0.044)	(0.044)	(0.021)	(0.025)
Bottom up/Control	0.164***	0.175***	0.199***	0.199***	0.185***	0.169***
	(0.018)	(0.019)	(0.044)	(0.044)	(0.022)	(0.026)
N	65426	57526	10827	10827	42074	31995
p	0.0	0.0	0.0	0.0	0.0	0.0
Log-likelihood	-132601.4	-116858.9	-22154.0	-22154.0	-85417.0	-65171.8
AIC	265230.7	233731.8	44322.1	44322.1	170848.0	130357.5

Notes: Each column shows a different model predicting the tax index dependent variable on different subsamples of countries. In each case, countries are excluded if they do not meet certain quality criteria. In all cases the effect sizes increase.

Table 6 shows that, using the combined Tax Index outcome, estimated treatment effects are remarkably stable when restricting the universe of cases considered, and if anything are larger when discarding observations from countries where treatment assignment predicts attrition. Table 11 and Table 12 in the appendix show this is also true of the two constituent

14

parts of the Tax Index measure: namely the Tax Morale outcome and the Tax Fine Attitude outcome.

Attrition and Lee trimming bounds

Lee trimming bounds describe the worst case scenario for how attrition could affect the results and therefore provide bounds for the size of the effect.[1] The control group had statistically lower attrition than the treatment groups in most countries in this study (84% of countries). However, this rarely seems to lead to any observable demographic imbalance in completed observations (12%). This is a mixed blessing because we lack pre-treatment covariates to narrow the Lee bounds (precisely because attrition is uncorrelated with observables). Since the Lee bounds do cross zero, we cannot mechanically dismiss the possibility that attrition partially explains some of our results (see table 7).

Nonetheless, the balance of evidence suggests that attrition is not driving our results. First, there is a very strong relationship between the effect sizes of the two treatments across countries (which is consistent with attrition), but only a very weak relationship between the effect size and attrition rate (which suggests that the country correlation is driven by other factors). The appendix provides further analysis on this point. Second, the lack of demographic imbalance post-treatment (including by taxpayer type) suggests that the attrition is not differentially affecting subgroups we can measure, which is how we would expect attrition to drive a treatment effect. Third, the robustness checks we run on subsets of countries that fail the various checks do not show smaller effect sizes (and in fact usually show larger ones). This is inconsistent with the results being driven by attrition. We therefore think the most likely cause of the differential attrition is simply that the control group was slightly less onerous than the treatment groups and that there was a general tendency among all subgroups exposed to the control condition to drop out of the survey at a slightly higher rate.

Table 7: Lee bounds analysis

	Combined				Tax fine				Tax morale			
	Top down		Bottom up		Top down		Bottom up		Top down		Bottom up	
	95% CI		95% CI		95% CI		95% CI		95% CI		95% CI	
Lower	-0.30	-0.19	-0.66	-0.58	-0.15	-0.09	-0.37	-0.30	-0.17	-0.11	-0.43	-0.39
Upper	0.64	0.73	0.86	0.96	0.33	0.39	0.54	0.60	0.31	0.36	0.46	0.50
Obs	174612		174439		174612		174439		174612		174439	
Selected obs	57899		54942		62638		59439		63100		59739	
Trimming prop	0.13		0.23		0.14		0.23		0.14		0.23	

Notes: Each column shows the results of Lee Bounds analysis showing the worst case scenario effect that differential attrition could have on the causal estimates from the experiment. The confidence intervals include zero meaning that we cannot definitively rule out the role of differential attrition.

---

[1] Note that the differential attrition only appears after the treatment is administered.

Mechanisms

Taken together, the above analyses strongly suggest that the top-down accountability treatment, and to a somewhat lesser extent the bottom-up participatory treatment, are effective at increasing tax morale among respondents.

Although inherently difficult to answer, one interesting question is whether one can identify beliefs that mediate this effect. With a view towards some initial answers, we asked respondents questions about whether abuse of money was common in their country, and whether politicians cared about what ordinary people needed. Our hypothesis was that both treatments might improve respondents' view of politicians, and that the top-down accountability treatment might reduce respondents' estimates of public sector corruption by increasing the salience of enforcement institutions and stressing the punishments they can inflict on wrongdoers.

However, these hypotheses fail to get any support in the data, as Table 8 shows. Both treatments fail to significantly affect beliefs about politicians, and both appear to increase respondents' beliefs about how widespread abuse of public money is. Many explanations for this are of course possible: for example, increasing the salience of public comptrollers might also simultaneously increase the salience of corruption. Alternatively, the mechanism survey instruments and/or their position at the very end of the survey might make them prone to measurement error, which might explain the null result on attitudes towards politicians. In any event, these results do not appear to account for why the treatments are effective at improving tax morale.

Table 8: Mechanism models (OLS)

	Politicians care	Abuse money
(Intercept)	0.343*** (0.003)	0.693*** (0.003)
Top down/Control	0.004 (0.004)	0.019*** (0.004)
Bottom up/Control	0.007 (0.005)	0.029*** (0.004)
N	65435	65435
p	0.3	0.0
Log-likelihood	-44238.4	-41293.9
AIC	88484.8	82595.9

Notes: Each column shows an OLS model predicting different variables capturing the possible experimental mechanisms of beliefs about politicians caring about ordinary people and politicians abusing money.

Treatment effect heterogeneity

Finally, we investigate how treatment effects vary based on covariates. In line with the theory of retributive justice, we had hypothesized in the pre-analysis plan that the top-down accountability treatment should have larger effects for "outsider" groups such as women, and lower for groups of taxpayers with larger direct exposure to the tax system

such as business owners. However, there is no support for this in the data: as Table 9 shows, there is very little evidence that business owners react differently from other respondents. In fact, we find no evidence that these treatments vary across taxpayer groups at all. This suggests that the treatment effects are not driven by a simple calculation about expected costs or benefits (which would vary across different types of taxpayers).

Table 9: Conditional Average Treatment Effects (OLS) by taxpayer type

	Tax index (OLS)	Tax fine (OLS)	Tax morale (OLS)
(Intercept)	5.029***	2.443***	2.586***
	-0.013	-0.008	-0.007
Top down/Control	0.239***	0.116***	0.123***
	-0.019	-0.011	-0.011
Bottom up/Control	0.166***	0.096***	0.070***
	-0.02	-0.012	-0.011
Business owner	0.08	0.077**	0.003
	-0.048	-0.028	-0.027
Public sector worker	0.273***	0.130***	0.143***
	-0.036	-0.021	-0.02
Top down/Control x Business owner	-0.066	-0.019	-0.047
	-0.069	-0.041	-0.039
Bottom up/Control x Business owner	-0.038	0.009	-0.047
	-0.071	-0.042	-0.04
Top down/Control x public	-0.036	-0.029	-0.007
	-0.052	-0.031	-0.03
Bottom up/Control x public	-0.054	-0.036	-0.018
	-0.054	-0.032	-0.031
N	65426	65429	65427
p	0	0	0
Log-likelihood	-133267.7	-98838.8	-96061.4
AIC	266555.4	197697.5	192142.7

Notes: Each column shows the results of an OLS model predicting each of the three dependent variables: tax fine, tax morale and the tax index summing the two. The treatments are interacted with public sector employment and being a business owner. None of the interactions show a significant effect.

Secondly, if anything, women appear to react much less strongly to the treatments (Table 10), and in particular to the bottom-up treatment, than men. Given the large literature on participatory institutions and women in politics, this is an interesting and provocative finding. However, one should of course point out that treatment effect heterogeneity cannot be interpreted causally, and there are many competing hypotheses here that we are unable to test given the available covariates.

Table 10: Conditional Average Treatment Effects (OLS) by gender

	Tax index (OLS)	Tax fine (OLS)	Tax morale (OLS)
(Intercept)	5.038***	2.456***	2.581***
	(0.014)	(0.008)	(0.008)
Top down/Control	0.254***	0.124***	0.130***
	(0.021)	(0.012)	(0.012)
Bottom up/Control	0.189***	0.108***	0.081***
	(0.021)	(0.013)	(0.012)
Female	0.101***	0.025	0.076***
	(0.026)	(0.015)	(0.015)
Top down/Control x female	-0.075*	-0.041	-0.034
	(0.038)	(0.022)	(0.021)
Bottom up/Control x female	-0.106**	-0.053*	-0.054*
	(0.039)	(0.023)	(0.022)
N	65429	65432	65430
p	0.0	0.0	0.0
Log-likelihood	-133329.6	-98882.4	-96112.6
AIC	266673.3	197778.9	192239.1

Notes: Each column shows an OLS model predicting the three dependent variables: tax fine, tax morale and the tax index summing the other two dependent variables. The experimental treatment are interacted with gender. The results show that the treatments are generally somewhat less effective for women.

Conclusion

Tax morale is an important phenomenon to understand, both for policy makers and academic researchers as well as the broader public. Here, we have documented experimental evidence from 50 countries about two interventions that, on average, generate significant increases in survey-based measure of tax morale: a participatory intervention allowing citizens to voice their expenditure preferences, and a top-down accountability intervention which increases the salience of anti-corruption efforts. We interpret our results as showing that even relatively short, shallow interventions can affect

beliefs and attitudes about tax morale substantially. In our view, this suggests that similar interventions could potentially be successful at increasing tax morale in real-world environments.

Some strengths and limitations of our study are inherent in the research design and method itself. On the positive side, the online survey technology employed here is relatively cost-effective (<1 USD per complete response), allowing us to scale the research design. On the other hand, due to the survey technology, we cannot gather many covariates, control attrition effectively or tap tax morale in a more direct, behavioral fashion. However, the characteristics of our research design are also related to two key contributions of the paper.

First, in our view, one important strength of the study is the extreme diversity of respondents in our sample: while previous studies especially on participatory interventions had typically been confined to very specific laboratory and political settings, we are able to document that our effects are relatively comparable, though by no means equal, across a very wide range of countries and contexts. Given that it is *ex ante* unclear how much homogeneity there is in responses to tax morale interventions, we believe that our results are a significant step forward.

A second key contribution we hope to make is to establish the effectiveness of "accountability" interventions that boost the salience of anti-corruption efforts. While previous research has hypothesized based on cross-country correlations that beliefs in the legitimacy of government generally, and to some extent the quality of democratic institutions, affect tax morale, we are able to show that experimentally varying information about anti-corruption efforts has measurable effects – by far the strongest in our study – on tax morale.

However, while we are able to show using manipulation checks that respondents actually absorb the information given to them in the interventions, our analysis finds no conclusive evidence on the question of why these interventions work to increase tax morale. We view this, as well as the question of how long-lasting these effects on tax morale are, as a fruitful avenue for further research.

Overall, our view is that the kind of highly-scalable research design documented here could help answer both important policy questions about tax morale, as well as social scientific questions about the social and psychological mechanisms underlying tax compliance around the globe. Especially when used in tandem with more context-specific studies featuring administrative data on tax compliance, this could constitute a useful part of a composite approach aimed both at generating and optimizing interventions and then deploying them in the field.

Appendix

Full survey

Intro	Your opinion is important to us. Thank you for your participation. Your answers will be kept anonymous.
Q1	Which of these best describes your situation?

Employee in the private sector
Employee in the public sector
Self-employed
Business owner, 1-24 employees Business owner, more than 24 employees
Unemployed and looking for work
Retired
Other

	Control Group	Top-down accountability intervention	Bottom-up participation intervention
Q2	Control - Q2A
There are many popular search engines in the world, with different designs and functions. Search engines are used every day by over 1 billion people worldwide.	Treatment 1 - Q2B
When government money is misused, it is very important to find and punish those responsible. Your government has a national agency, the _____, that helps to punish the misuse of government funds.	Treatment 2 - Q2C
You have been selected to be part of the Online Citizen Assembly: a national conversation about how the government of _____ should spend money. The results of the Online Citizen Assembly will be presented to the government.
	Continue	Continue	Continue
Q3	Q3A
Many people say they are annoyed by all the advertising on search engines. How much does this apply to you?	Q3B
The _____ has investigated many cases of government corruption. Many people who misused government funds have been punished. Do you think it is good to have an agency that investigates government corruption?	Q3C
What should your government spend more money on? The results of the Online Citizen Assembly will be presented to the government.
	Not at all	Yes	Defense and Police

A bit	No	Education
Quite a bit		Transportation
Very much		Welfare
		Health
		Environment

Q4   If a taxpayer does not report all of their income in order to pay less income taxes, what percentage of their income should they pay as a penalty?

None (0%)
1-10%
11-20%
More than 20%

Q5   If a taxpayer does not report all of their income in order to pay less income taxes do you feel it is:
Not wrong
A bit wrong
Wrong
Seriously wrong

Q6   [manipulation check A] What type of corruption does the [comptroller name here] investigate?
Government
Private sector
Not sure

Additional questions:

Q7   [manipulation check B] After taking this survey, do you feel like you have opportunities to tell the government how to spend money?
Yes
No

Q8   Do you think that politicians care about what ordinary people need?

Yes
No

Q9   How common is abuse of public money in [country name here] ?

Common
Not common

Outro Thank you for participating in this study of accountability and government spending. This study is part of a research project. Your answers are anonymous and the combined results may be used in scientific publications and presentations.

Correlation of country effects and retention

If differential retention is responsible for the treatment effects in this study, we would expect the effects to be stronger in countries with higher differential retention. Because our estimates of both of these effects are relatively noisy, we use a random slopes/random intercepts model to pool the variance towards the group means. For the experimental effects, that leads to us estimating a strong linear relationship between the top down and bottom up effect sizes (see figure 5).
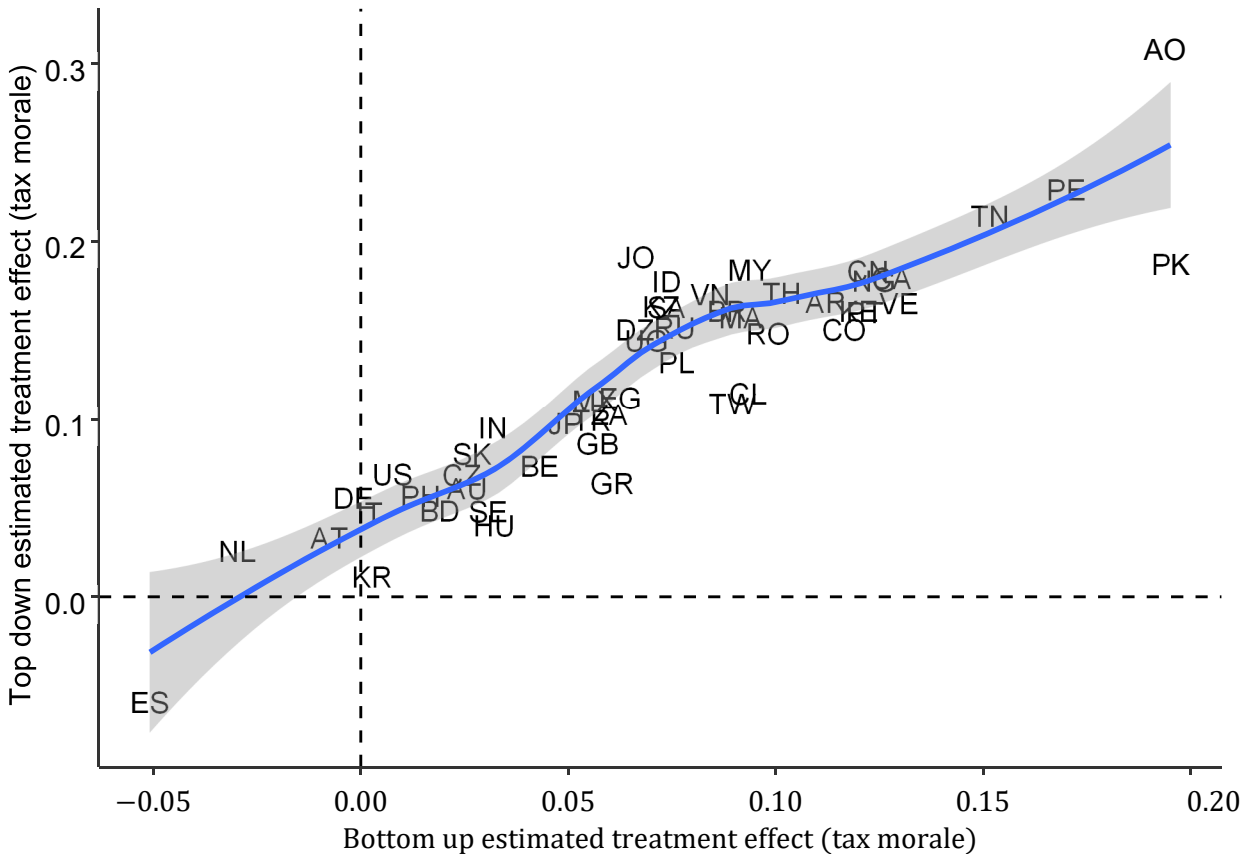

Figure 5: Estimated country effects of top down and bottom up treatment groups on GSS tax morale question (estimated effect=random slope + fixed effect slope. loess regression used to plot smoothed line).

One possible explanation for the strong relationship between the estimated effects is that both effects are driven by differential attrition (which would boost both treatments compared to the control group). However, if this was the case, we should also see a very strong relationship between estimated treatment effects on retention and the estimated treatment effects on tax morale. We use a second random intercepts/random slopes model to estimate the treatment effects on retention across countries. We then plot these estimated intercepts against the estimated treatment effects (Figure 6).
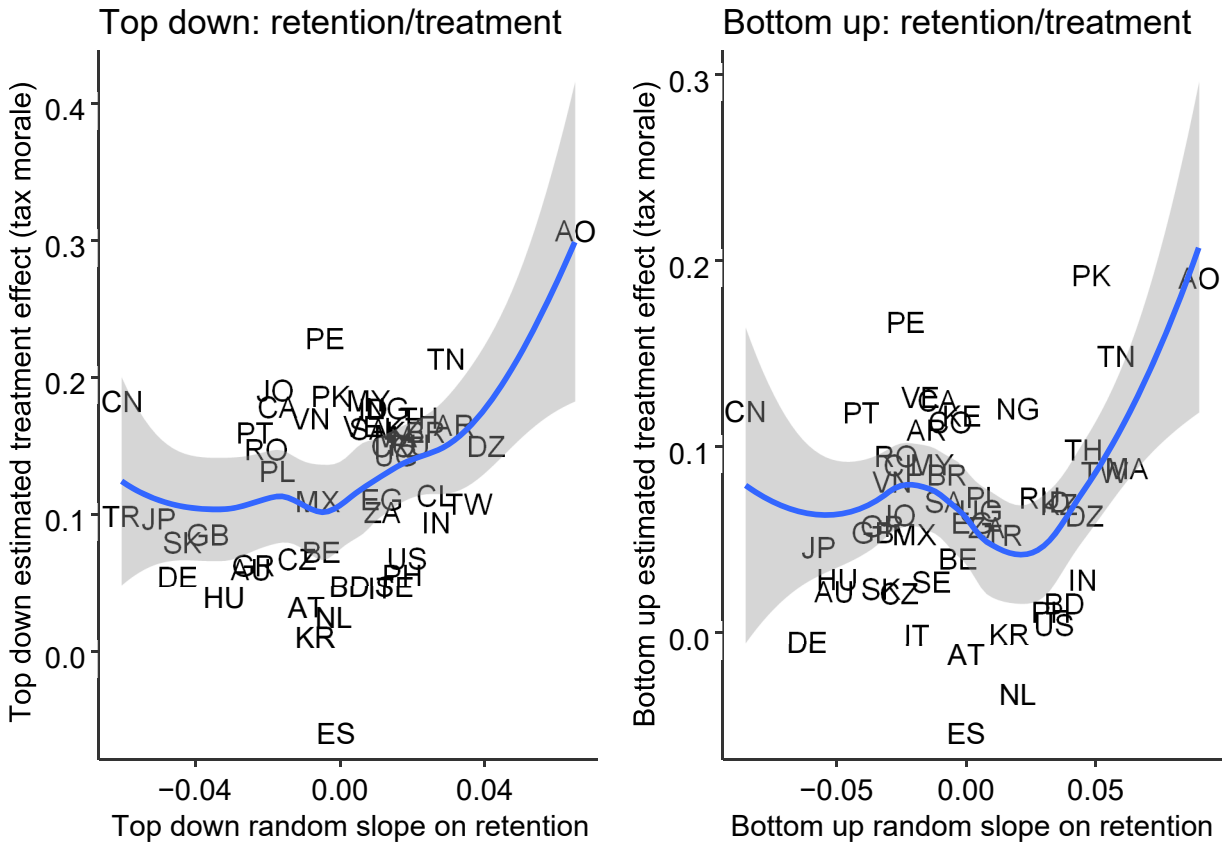
Figure 6: Random slopes of treatment group on retention rates plotted against estimated country effect
(random slope + fixed effect slope) for top down and bottom up treatments (loess regression used to plot smoothed line).

The relationship is clearly weak and non-linear and is largely driven by the outlier of Angola. Overall, the strength of the estimated relationship does not appear sufficiently strong to indicate that differences in attrition explain the differences in the size of the treatment effects across countries. This also provides another piece of evidence against differential attrition being the key driver of our experimental results.

Country means

The raw means for each country and experimental group are shown in Figure 7.
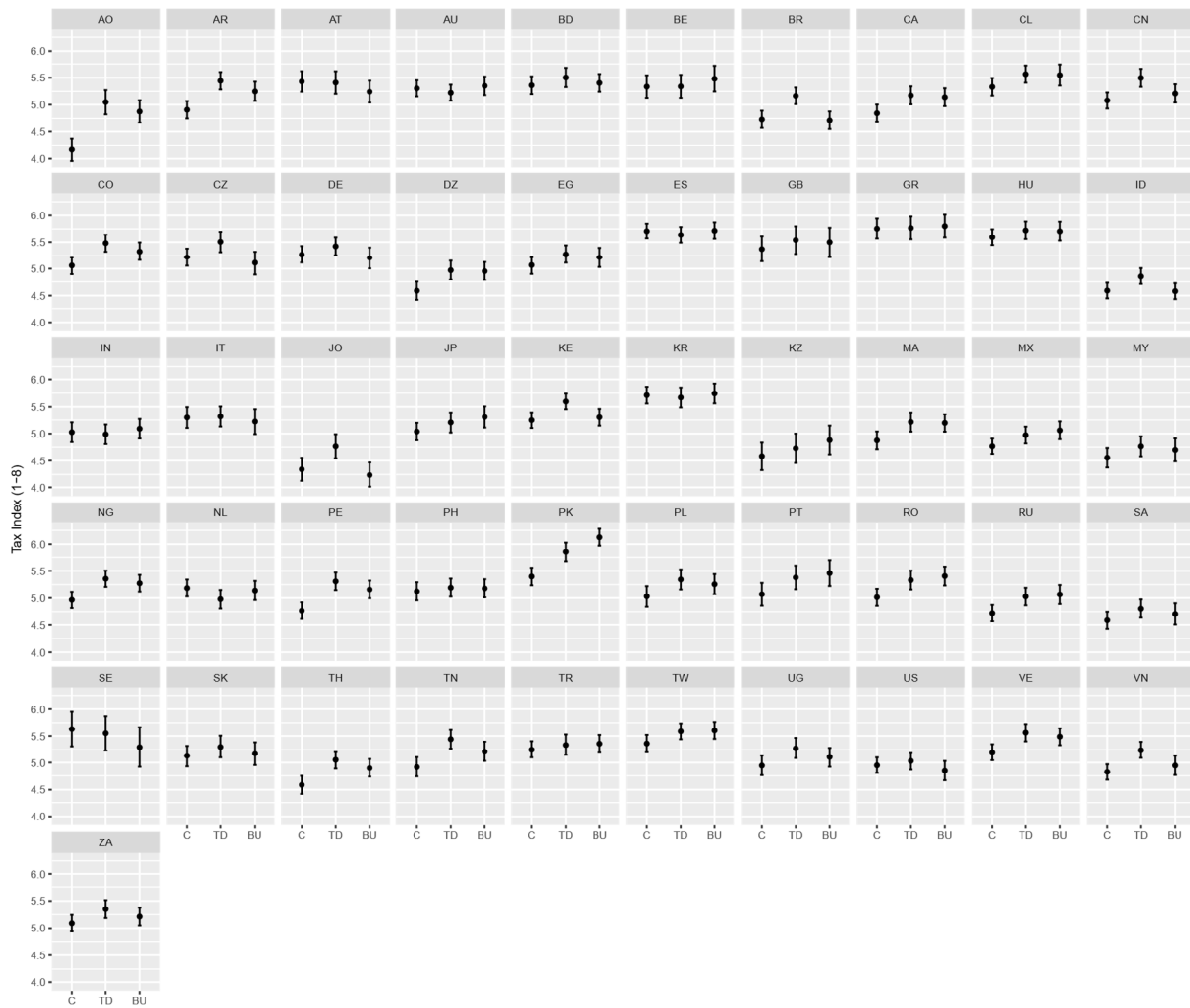

Figure 7: Simple comparison of means, by country, Tax Index (sum of outcomes) (95% CIs)

Robustness Checks for Constituent Dependent Variables

The robustness checks for each constituent question of the tax morale index show the same pattern of results as for the combined index.

Table 11: Main Model and Robustness Checks: Tax Morale (OLS)

	Main	Balance	Attrition	Attrition (cond)	Bottom manip	Top manip
Top down/Control	0.120***	0.118***	0.131***	0.131***	0.138***	0.114***
	(0.010)	(0.010)	(0.025)	(0.025)	(0.012)	(0.014)
Bottom up/Control	0.069***	0.068***	0.095***	0.095***	0.079***	0.075***
	(0.010)	(0.011)	(0.025)	(0.025)	(0.013)	(0.015)
N	65427	57526	10826	10826	42073	31993
p	0.0	0.0	0.0	0.0	0.0	0.0
Log-likelihood	-95363.8	-83719.9	-16136.1	-16136.1	-61503.1	-47196.3
AIC	190755.5	167467.8	32300.2	32300.2	123034.3	94420.7

Table 12: Main Model and Robustness Checks: Tax Fine Attitude (OLS)

	Main	Balance	Attrition	Attrition (cond)	Bottom manip	Top manip
Top down/Control	0.111***	0.117***	0.146***	0.146***	0.125***	0.108***
	(0.010)	(0.011)	(0.026)	(0.026)	(0.013)	(0.015)
Bottom up/Control	0.095***	0.105***	0.107***	0.107***	0.104***	0.093***
	(0.010)	(0.011)	(0.026)	(0.026)	(0.013)	(0.015)
N	65429	57528	10828	10828	42074	31994
p	0.0	0.0	0.0	0.0	0.0	0.0
Log-likelihood	-98487.5	-86770.6	-16483.2	-16483.2	-63393.9	-48430.2
AIC	197003.1	173569.2	32994.5	32994.5	126815.9	96888.4

References

Ali, Merima, Odd-Helge Fjeldstad, and Ingrid Hoem Sjursen. 2014. "To Pay or Not to Pay? Citizens' Attitudes Toward Taxation in Kenya, Tanzania, Uganda, and South Africa." *World Development* 64. Elsevier: 828–42.

Allingham, Michael G, and Agnar Sandmo. 1972. "Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics* 1 (3-4). Elsevier: 323–38.

Alm, James, and Benno Torgler. 2006. "Culture Differences and Tax Morale in the United States and in Europe." *Journal of Economic Psychology* 27 (2). Elsevier: 224–46.

Alm, James, Betty R Jackson, and Michael McKee. 1992. "Estimating the Determinants of Taxpayer Compliance with Experimental Data." *National Tax Journal.* JSTOR, 107–14.

Alm, James, Gary H McClelland, and William D Schulze. 1992. "Why Do People Pay Taxes?" *Journal of Public Economics* 48 (1). Elsevier: 21–38.

Casal, Sandro, Christoph Kogler, Luigi Mittone, and Erich Kirchler. 2016. "Tax Compliance Depends on Voice of Taxpayers." *Journal of Economic Psychology* 56. Elsevier: 141–50.

Castro, Lucio, and Carlos Scartascini. 2015. "Tax Compliance and Enforcement in the Pampas Evidence from a Field Experiment." *Journal of Economic Behavior & Organization* 116. Elsevier: 65–82. Daude, Christian, Hamlet Gutiérrez, and Ángel Melguizo. 2012. "What Drives Tax Morale?" OECD Publishing.

Del Carpio, Lucia. 2013. "Are the Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru." *Work. Pap., Princeton Univ., Princeton, NJ.*

Dwenger, Nadja, Henrik Kleven, Imran Rasul, and Johannes Rincke. 2016. "Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany." *American Economic Journal: Economic Policy* 8 (3). American Economic Association: 203–32.

D'Arcy, Michelle. 2009. "Why Do Citizens Assent to Pay Tax? Legitimacy, Taxation and the African State?" Feld, Lars P, and Jean-Robert Tyran. 2002. "Tax Evasion and Voting: An Experimental Analysis." *Kyklos* 55 (2). Wiley Online Library: 197–221.

Fjeldstad, Odd-Helge, Collette Schulz-Herzenberg, and Ingrid Hoem Sjursen. 2012. "People's Views of Taxation in Africa: A Review of Research on Determinants of Tax Compliance."

Flores-Macías, Gustavo A. 2016. "Building Support for Taxation in Developing Countries: Experimental Evidence from Mexico." Institute of Development Studies.

Gobena, Lemessa Bayissa, and Marius Van Dijke. 2016. "Power, Justice, and Trust: A Moderated Mediation Analysis of Tax Compliance Among Ethiopian Business Owners." *Journal of Economic Psychology* 52. Elsevier: 24–37.

Kettle, Stewart, Marco Hernandez, Michael Sanders, Oliver Hauser, and Simon Ruda. 2017. "Failure to Captcha Attention: Null Results from an Honesty Priming Experiment in Guatemala." *Behavioral Sciences* 7 (2). Multidisciplinary Digital Publishing Institute: 28.

Lamberton, Cait, Jan-Emmanuel De Neve, and Michael. Norton, 2014. "Eliciting Taxpayer Preferences Increases Tax Compliance."' SSRN Working Paper

Lue, Xiaobo, and Lily. Tsai., 2017.'Citizen Input and Preferences for Tax Compliance.' Manuscript.

Luttmer, Erzo FP, and Monica Singhal. 2014. "Tax Morale." *The Journal of Economic Perspectives* 28 (4). American Economic Association: 149–68.

Ortega, Daniel, Lucas Ronconi, and Pablo Sanguinetti. 2016. "Reciprocity and Willingness to Pay Taxes:

Evidence from a Survey Experiment in Latin America." *Economía* 16 (2). Brookings Institution Press: 55–87.

Pommerehne, Werner W, and Hannelore Weck-Hannemann. 1996. "Tax Rates, Tax Administration and Income Tax Evasion in Switzerland." *Public Choice* 88 (1). Springer: 161–70.

Slemrod, Joel, and Shlomo Yitzhaki. 2002. "Tax Avoidance, Evasion, and Administration." *Handbook of Public Economics* 3. Elsevier: 1423–70.

Torgler, Benno. 2005. "Tax Morale and Direct Democracy." *European Journal of Political Economy* 21 (2). Elsevier: 525–31.

Tsai, Lily L, Minh Trinh, Shiyao Liu. 2017. "The Role of Retributive Justice in Citizen Evaluations of Government: The Case of China". MIT Political Science department Research Paper No. 2017-24

Tsai, Lily L, 2019, *When People Want Punishment: Retributive Justice and the Puzzle of Authoritarian Popularity*, Book manuscript.

Yesegat, Wollela Abehodie, and Odd-Helge Fjeldstad. 2016. "Business People's Views of Paying Taxes in Ethiopia." Institute of Development Studies.

Xiaobo, Lü, Lily L. Tsai, and Ping Zhang. 2018. "Procedural Justice, Retributive Justice, and Property Tax Compliance: Experimental Evidence from China". Conference Paper, Annual Meeting of the American Political Science Association, Boston, MA.