PRE-ANALYSIS PLAN

"Positive vs. Negative Incentives for Compliance: Evaluating a Randomized Tax Holiday"

Thad Dunning,[*] Felipe Monestier[†] Rafael Piñeiro,[‡]
Fernando Rosenblatt,[§] and Guadalupe Tuñón[¶]

July 23, 2014

This pre-analysis plan is posted to the study registries of the American Economic Association as well as the Experiments in Governance and Politics (EGAP) network. The study has been approved by the Office for the Protection of Human Subjects at the University of California, Berkeley (Protocol ID 2014-04-6286). We are grateful to Lihuen Nocetto for research assistance and to the J-PAL Governance Initiative for generous funding.

[*]University of California, Berkeley.
[†]Universidad de la República, Montevideo, Uruguay
[‡]Universidad Católica, Montevideo, Uruguay
[§]Universidad Diego Portales, Santiago, Chile
[¶]University of California, Berkeley

Abstract

Can positive rather than negative incentives boost tax compliance in developing countries? We study a unique randomized policy innovation in Montevideo, Uruguay, in which the municipal government raffles tax holidays to good taxpayers who are current on past payments. Using unusual access to over-time tax payment records as well as survey data, we assess the impact of holidays on subsequent tax compliance, as well as citizens' attitudes towards taxation and governance. We also use field and survey experiments to study the effects of informing eligible and ineligible taxpayers about the rebate lottery—which has not been effectively advertised by the government. Our informational treatments allow us to compare the influence of priming negative incentives for tax compliance, such as fines and punishment for non-payment, with the positive inducement provided by the lottery.

Keywords: Tax compliance, developing countries, state capacity, positive vs. negative incentives; field experiment, natural experiment, information, lottery

Contents

List of Figures

List of Tables

1 Introduction

Developing countries face enduring obstacles to engendering tax revenue, a key facet of state capacity. To explain these obstacles, researchers often focus on the difficulty of sanctioning non-payment of taxes. According to this reasoning, states can boost tax compliance using *negative* incentives, such as fines and other punishments for non-payment; however, enforcement is difficult for states with poor administrative capacity. Thus, the extent of compliance is a function not only of the severity of sanctions but also the probability of their application. In this view, weak state institutions are responsible for persistent non-compliance in developing countries.

Yet, governments have occasionally used *positive* rather than negative incentives to generate tax compliance.[1] In this project, we study an unusual policy innovation in Montevideo, Uruguay, in which tax holidays are randomly assigned to eligible taxpayers. Since 2004, and across four kinds of taxes (property, vehicle, sewage, and head), the municipality of Montevideo has randomly selected taxpayers and—conditional on a recent history of good taxpaying—rewarded them with a year free of tax payments. Rather than being punished for poor tax-paying behavior, citizens are therefore rewarded for their history of tax compliance.

Insights from psychology and behavioral economics suggest that promised rewards may have substantially different effects than threatened punishments. Yet the impact of positive incentives for tax compliance has not to our knowledge been reliably assessed. Nor have researchers systematically compared the effectiveness of positive and negative encouragements to pay taxes. Indeed, given severe selection problems in non-randomized observational studies, it is typically challenging to answer questions about the effects of positive and negative inducements.

In the research we register with this document, we use the natural experiment provided by Montevideo's randomized tax holiday, and our unusual access to a panel of administrative tax payment records, to study the impact of winning a tax holiday. The lottery sets up a straightforward comparison, among eligible taxpayers—those with good tax-payment records—between winners and a randomly selected control group of eligible non-winners. Thus, we can assess whether winning the lottery, which not only provides a year free of tax payments but also informs many taxpayers of the existence of the program,

[1] For example, countries such as Argentina and Uruguay have raffled prizes to taxpayers who turn in receipts for value-added taxes. The objective of such policies is to increase the reporting of sales transactions by vendors.

influences subsequent tax compliance. Varying the period of elapsed time between winning the lottery and outcome measurement allows us to assess the persistence of effects on tax compliance. We also use household survey data for a sub-sample of winners and eligible non-winners to compare attitudes towards the equity and fairness of the tax system, as well as broader political attitudes, support for political incumbents, and so forth. For the natural experiment, we focus on the effects of winning any of the four types of taxes subject to tax holidays, as we have administrative tax payment data available for all four types of taxes.

However, the impact of the program is likely greater than these comparisons will suggest, because knowledge of the program may influence citizens to bring their tax accounts up to date (to become good taxpayers) in order to gain eligibility for tax holidays. We therefore also utilize a supplementary field experiment to assess whether informing citizens, including bad taxpayers, about the existence of the lottery affects subsequent tax compliance. In particular, in collaboration with the municipal government, we mail flyers stamped with the municipal logo—which appear very much like tax bills themselves—reminding taxpayers of the due date for taxes and providing experimentally varied messages to taxpayers. For these informational interventions, we focus on property (real estate) taxes, as these are the most visible and important of the four taxes we study.

In more detail, we experimentally vary the content of messages to study the effects of negative and positive inducements for tax compliance. Thus, we compare taxpayers who are 1. simply reminded of an impending due date for real estate taxes (placebo control group) to: 2. those who are reminded/informed of the existence of the tax holiday lottery and 3. those who are reminded/informed of the existence of punishments for non-payment of taxes.[2] For group 2., we further subdivide the treatment group into those who receive 2A. no information about the probability of winning the lottery and 2B. those who are told their probability of winning (which we compare to taxpayers' priors about the likelihood of winning, estimated from follow-up survey data on the control group). Finally, in addition to these treatments priming individual rewards and punishments for payment or non-payment of taxes, we use alternative interventions that additional prime either 4. the social rationale for the lottery or 5. the social rationale for punishment of non-compliance with taxes. Our social treatments may manipulate the individual normative/expressive benefits of paying taxes and thus may influence

[2] See subsection 3.2.1 for the text of the informational interventions.

compliance in distinct ways. The results of our informational experiment will allow us to compare the impact of manipulating negative and positive incentives for tax payment, thus shedding light on basic motivations for tax compliance, and also evaluate the best means of boosting future impact through informational interventions that effectively advertise the policy. The findings may therefore suggest how to maximize impact when the program is scaled up or transported to other settings.

Our study has several key advantages. First and foremost, the randomization of the natural and field experimental interventions sets up straightforward, unconfounded comparisons between the treatment groups, aiding inferences about the causal impact of the interventions. Second, we make use of detailed, unobtrusive outcome measures—namely, individualized administrative records on tax compliance, extent of tax debt, and so forth. These records are akin to turnout data in voter mobilization experiments, but unusually they allow unobtrusive measurement of *tax payment*. Third, the experimental realism of our interventions is quite high, since we designed flyers that match the format of tax bills and arrive bearing the municipal logo—just as would real interventions designed to boost compliance by publicizing negative or positive incentives for tax payment.

Finally, our factorial experimental design and supplementary measurement using household surveys will shed tentative light on the mechanisms through which positive and negative inducements generate compliance. For example, winning the lottery might shape compliance by boosting income or altering habits (which could conceivably lead to a negative impact on compliance), by notifying uniformed taxpayers of the existence of the lottery, or by shaping beliefs about the equity and transparency of the tax system as well as broader political attitudes. Though distinguishing such mechanisms is tricky, as a recent literature on causal inference highlights, we use several strategies to explore possible *reasons* for the impact of promised benefits or threatened punishments. For example, we compare the effect of winning a tax holiday lottery to the effect of merely informing eligible non-winners about it (and use data on the proportion of uninformed taxpayers from our household surveys to generate instrumental-variables estimators). We also test for heterogeneous effects that shed light on the impact of rewards, for example, we assess whether effects depend on the size of past tax arrears, on the theory that manipulating material incentives for payment are greatest for "taxpayers at risk" (e.g., good taxpayers who have not always been good, or bad taxpayers who have not always been bad). We also use survey as well as field and natural experiments to compare the effects of negative and positive inducements on beliefs about the tax system as well as broader political attitudes; effects on such outcomes are of interest because beliefs and attitudes may in turn influence compliance, but also because they represent independently interesting and important consequences of the tax holiday lottery.

In the rest of this document, we sketch the theory that motivates our study, particularly our comparison of negative vs. positive incentives for tax compliance (Section 2); describe our empirical strategy, including our natural and field experimental designs and the timing and nature of data collection (Section 3); discuss our outcome measures (Section 4); and outline our hypotheses and statistical tests, including adjustments for multiple comparisons (Section 5).

2 Theory

2.1 What hinders tax compliance?

Engendering tax compliance is a routine problem, especially in developing countries. According to most accounts, states with weak administrative capacity fail to "penetrate" society sufficiently, thereby allowing citizens to avoid payment of taxes by not imposing credible penalties for non-payment. Thus, scholars typically see information and monitoring problems at the heart of the problem, in which lack of state capacity explains failures to elicit compliance.

This account relies on an underlying behavioral theory: the decision not to pay taxes is driven by the benefit of evasion, minus the cost of punishment discounted by the probability of detection. The problem, according to this theory, is thus that in developing countries the probability of punishment for non-compliance is negligible. As a corollary, collecting taxes may be politically unpopular, which may provide an additional incentive for elected governments to opt not to enforce penalties for non-payment of taxes. In sum, failure to generate tax revenue is seen as a problem of enforcement that is due in part to weak state capacity.

Yet, in developed and even developing countries alike, many people do pay taxes—even when penalties and chances of punishment for evasion are quite low. This is puzzling from the perspective of a focus on negative incentives. Indeed, the simple decision framework we develop in this section suggests that if compliance were only driven by the benefit of evasion minus the expected cost of punishment, nobody would pay taxes at all. Some lab experimental evidence also suggests that compliance behavior can be insensitive to the cost of penalties or the probability of punishment. What then explains why in many settings people *do* pay taxes?

We investigate in this study the impact of *benefits* for tax compliance—in particular, the positive incentive provided by the randomized tax holiday for good taxpayers in Montevideo. Individual benefits play a key role in theories of voting over tax policy, but the implications for compliance are underexplored. Note that benefits of paying taxes may be *material* (e.g. a direct individual or group return to taxes, in the form of public spending) or *expressive/normative*, e.g., due to social preferences, equity considerations, altruism, or burden sharing. A tax holiday lottery such as the one we study might shape both the perceived material and expressive benefits of paying taxes; moreover, different ways of framing the existence of the lottery (for instance, emphasizing its social rationale) might influent the individual expressive/normative benefits of payment. The impact of this lottery has not been widely studied, and thus our research question is empirical: do individual rewards influence tax *compliance* behavior, and if so, why?

Note that the tax setting we study is somewhat different than in the classic compliance problem. Here we study municipal real estate, vehicle, sewage, and head taxes, in which the tax due is known by the taxing authority, e.g. because the value of assets is appraised). Thus, information and monitoring problems are plausibly less severe than in many settings, such as the payment of income tax.[3] Nonetheless, enforcement remains a central issue in this context: the municipal government can decide to pursue delinquent taxpayers more or less aggressively—sometimes failing to pursue bad taxpayers for a period of many years and sometimes using the courts to expropriate the property of bad taxpayers—and the outcome of any individual renegotiation is ex-ante uncertain for taxpayers. As we discuss below, Montevideo's city government has frequently issued general amnesties for bad tax payers, and it often renegotiates debt with individual taxpayers as well. For this reason, punishment for non-compliance is very far from certain, which tends to make tax payment mysterious from the point of view of the classic decision calculus we sketch next—in which the choice not to pay taxes is driven by the benefit of non-payment or delay in payment, minus the expected cost of punishment. Our set-

[3] In this sense, the problem facing the state is akin to that of a credit-card company faced with non-payment by consumers: the value of debt is known, and hiding the amount owed is not an option, but the company can still face substantial barriers to collecting taxes.

ting therefore provides a useful vehicle for comparing the impact of positive vs. negative incentives for tax payment: in particular, how the probability of punishment and the possibility of rewards condition tax compliance; and how manipulation of other factors—such as the *benefits* of paying taxes—may engender greater tax compliance.

2.2 The role of positive incentives

A simple formalization of the decision problem highlights the difficulty of encouraging tax compliance. Let y be an asset value, t be the annual tax rate, and z be the unpaid annual amount of taxes due; with full nonpayment, $z = ty$. The expected utility of full nonpayment in any year is thus

$$z - pc, \tag{1}$$

where c is the penalty for nonpayment and p the probability of punishment. In the setting we study, the cost of punishment c could include (1) fines and interest charges for delayed payments, and ultimately (2) losing one's house or other property.[4] However, a taking of property by the city government is not certain, even for the worst taxpayers, given the legal costs involved. Moreover, the possibility of a future amnesty for bad taxpayers, which happen frequently as we discuss below, means that the probability of punishment p is very far from one in this context.[5]. The fine for nonpayment c is often small.

Under the policy we study, good taxpayers win a year free of tax payments with probability 1/5,000 in any tax payment period. Thus, the expected utility of paying the full year's taxes this year is (without discounting)

$$\frac{1}{5,000}z - z. \tag{2}$$

Note that the direct material benefit provided by the lottery is limited, relative to the cost of compliance here: in order to gain eligibility for the lottery, one has to pay a year's worth of taxes z, while gaining a year free of tax payments in the following year only with probability $1/5,000$. Indeed, comparing (2)

[4]Property owners also cannot sell their houses until clearing their property tax accounts.
[5]Note that $(1 - p)$ is the (unknown) probability of amnesties and individual renegotiations

with (1), we see that paying taxes is only optimal when

$$pc > \frac{9,999}{5,000}z. \tag{3}$$

In words, the cost of the penalty discounted by the probability of punishment must be essentially as large as two full years of taxes due on the asset.

A natural question that arises is: does manipulation of the expressive or perceived benefit of paying taxes further influence compliance? Let b be the perceived benefit of paying taxes, in terms of the social return or legitimacy of taxation. The expected utility of paying taxes is then $(1/5,000)z - z + b$, so tax payment occurs whenever

$$pc > \frac{9,999}{5,000}z - b, \tag{4}$$

which is satisfied more easily than is (3). Depending on the size of b—the perceived benefit of paying taxes—it could be incentive compatible for taxpayers to comply, even when pc is small.

Our theory is thus that beyond providing a direct benefit—which is vanishly small in expectation—Montevideo's lottery may also affect b, e.g. by influencing perceptions of the fairness and equity of taxes and/or the individual normative benefit of paying taxes. In some sense, this must be true if the lottery policy affects compliance, at least if people make payment decisions according to (4). Among property owners, the average annual value of property taxes is over US\$1,000 (24,000 Uruguayan pesos), which is non-trivial[6]; but the *expected* value of the lottery US\$1,000/5,000, or about twenty US cents. We therefore expect the material benefits of the lottery to influence compliance only if (1) people misperceive the probability of winning the lottery or (2) the benefits provided by the lottery exceed the expected material payoff, e.g. because rewards influence behavior differently from punishments, or because the transparency and legitimacy of the lottery policy influence expressive benefits b.

This decision framework also suggests that manipulating benefits may have a bigger effect on certain kinds of taxpayers, in terms of pushing them over the threshold to compliance:

- Taxpayers who owe less (e.g. for whom the amount owed $z < ty$);
- Those with higher income or capacity to pay, relative to asset values; and

[6]This is an estimate from an interview with the IM; we will verify this using our probability sample of tax records.

- Those with high subjective p, i.e., who overestimate the probability of winning the lottery).

We discuss these hypotheses about heterogeneous effects for such marginal taxpayers—those with small amounts of indebtedness or imperfect but not terrible records of compliance—further in Section 5.

3 Empirical Strategy and Design

The innovative tax holiday policy developed by Montevideo's City Hall (Intendencia de Montevideo—IM) seeks to improve tax compliance by providing positive incentives for good taxpaying; it also counteracts negative perceptions among citizens of forgiveness for non-compliance. The policy was initiated by the center-left government of the Frente Amplio in the context of an amnesty for many delinquent taxpayers following the economic crisis of 2002. The idea was to counteract perceived negative incentives of the tax amnesty.[7] As officials at the IM have told us, the economic crisis generated a dilemma: how to lower the burden for those under dire circumstances while at the same time continuing to promote compliance. The lotteries were their answer.[8] After almost ten years, however, no evaluation of the program's effectiveness exists.

To select taxpayers for holidays, the government uses the results of Uruguay's National Lottery, which posts online five random digits that indicate winning lottery numbers. Taxpayers whose four-digit IDs correspond to the final four digits of the winning National Lottery numbers are selected as the provisional winners of tax holidays.[9] The National Lottery frequently posts lottery results, which the municipal government uses to select winners of rebates for each type of tax. The municipality sends a letter to eligible winners whose tax accounts are paid up in the previous fiscal year (i.e., "good" taxpayers) indicating that they should register for a year free of tax payments; registration allows the

[7] For discussion of the initiation of the lottery, see http://historico.elpais.com.uy/Suple/LaSemanaEnElPais/04/10/29/lasem$_c iud_1$18264.asp; htt $//www.montevideo.com.uy/notnoticias_6$6228$_1$.$html$; $andhttp$: $//www.$180.$com.uy/articulo/$14284.

[8] In October 2013, the municipality announced a renewed amnesty for certain bad taxpayers, underscoring the difficulties of cracking down on non-compliance. There have been amnesties in 2004, 2008, and 2010, among other years. See http://www.montevideo.com.uy/notnoticias$_6$6228$_1$.$htmlandhttp$: $//www.$180.$com.uy/articulo/$14284.

[9] The randomization occurs through the selection from balls from an urn, as described in Spanish at http://www.loteria.gub.uy/Juego_Loteria.php. For an example of posted lottery results, see http://www.loteria.gub.uy/ver_resultados.php?vdia=21vmes=6vano=2013. Winning taxpayer numbers are posted at http://www.montevideo.gub.uy/sorteosBP/pages/sorteosBuenosPagadores.xhtml.

city government to screen winners to ensure that they are physical persons and not, e.g., corporations. For good taxpayers, the probability of winning any given rebate lottery is 1/10,000; however, some taxpayers who would have won a tax lottery are not good taxpayers or physical persons or do not present themselves to the city government after being notified. Also, the municipality grants such holidays six times a year (February, April, June, August, October, and December) for head and sewage taxes, and three times a year (March, July, and November) for vehicle and property (real estate) tax. Because the real estate and vehicle taxes are paid three times a year, with two lotteries in the interim (and winning either lottery grants a year free of tax payments), for those taxes the probability of winning a year free of taxes in any tax period is 1/5,000. We focus in the project mainly on the impact of the tax holiday for property taxes, though we also estimate effects of winning the lottery for the other three taxes.

We will use both administrative data (tax payment records) provided by the municipality and household survey data (both described further in Section 4) to study the effects of the lottery as well as the effects of our informational interventions. Due to the greater cost involved in data collection for the household surveys, our household survey will be administered to a random sample of individuals for whom we have administrative/tax payment data. Tables 3.1 and 3.2 in the next sub-sections give the sample sizes for each of the two datasets, which we justify using power calculations in Section 3.4. In the next subsections, we describe how we will use comparisons between the cells of Tables 3.1 and 3.2 to estimate different kinds of causal effects.

In more detail, we asked the municipality for taxpayer records for all eligible lottery winners since 2004 (we currently believe that is circa 7,200 taxpayers but may be fewer), and a random sample of eligible ("good") taxpayers who have not won the lottery of approximately the same size (see Table 3.1)); below, we describe the sampling procedure for the latter group. These samples comprise the treatment and control groups in the natural experiment. For the study group for the field experiment, we asked for a random sample of eligible ("good") taxpayers (N=14,250) and a random sample of ineligible ("bad") taxpayers (N=14,250) (see Table 3.2). We then draw a sub-sample of size 8,000 (N=2,000 from the natural experimental study group, N=6,000 from the field experimental study group) for household surveys. The administrative data list taxpayer names, addresses and in some cases phone numbers; our survey firm will use those data to track down sampled individuals. In cases where sampled individuals cannot be found, are not physical persons, or refuse to participate, we sample replacement households at random from the taxpayer records included in the study group. The survey will be conducted by the Uruguayan firm CIFRA.

3.1 Natural experiment: the effect of winning a tax holiday

The design of the lottery allows us readily to estimate the effects of winning a tax holiday, among good taxpayers. In particular, we will use a time-series panel of administrative data (2004-2013) to assess the effects of winning the lottery on subsequent tax payments, comparing the payment history at $t + 1$, $t + 2$, $t + 3 \ldots$ of lottery winners to a control group of eligible non-winners, where t is the year (or portion of year) in which winners won the lottery.

Constructing appropriate treatment and especially control groups requires some care. For the treatment group, we will have access to tax payment data for all winners of each tax lottery since 2004. We will check to ensure that the last four digits of winners' taxpayer IDs (current account numbers) match the last four digits of the winning National Lottery number in the corresponding lottery. Municipal officials estimate the number of winners of tax holidays across all lotteries at 7,200 (though we believe the number may be smaller); we will receive the data on approximately July 25, 2014, shortly after filing this pre-analysis plan (see sub-section 3.3.1). We will request data for both eligible and ineligible taxpayers whose IDs correspond to the four-digit numbers of lottery winners.[10]

For the control group, note first that the right counterfactual group for winners of a particular lottery are taxpayers who were eligible to win as of the date of *that* lottery, based on being current on their tax payments over the previous year.[11] To select a random sample of eligible lottery non-winners for *each* lottery since 2004, we randomly generate a single four-digit number for each lottery since 2004, replacing at random any numbers that coincide with the IDs of lottery winners.[12] The municipality will then locate all taxpayer IDs that end in each random four-digit number and identify those who were eligible to win the corresponding lottery, based on their tax compliance status at the time the

[10]The treatment and control groups are comprised only of eligible taxpayers, but data on ineligible taxpayers will be useful for a number of purposes, such as placebo tests

[11]If we used a control group of currently eligible taxpayers, we would risk bias: the treatment group would include only taxpayers who were eligible to win (good taxpayers) as of the date of each lottery, while the control group would include a mix of taxpayers who were eligible and who were ineligible as of the date of each lottery.

[12]These data will be provided to us free of charge, yet recovering specific taxpayer records requires effort from the municipal bureaucracy; thus, we cannot use a census of administrative data in the control group, as we do in the treatment group.

lottery took place. Thus, our random procedure for constructing the control group exactly mimics the random process that created the treatment group of lottery winners. The average number of winners per lottery is approximately 20, and thus so is the average size of our control group for each lottery, given the manner in which we construct it. Both the treatment and control groups are comprised of random samples from the population of good taxpayers in Montevideo, as of the date of each lottery.

Table 3.1: **Natural Experiment: Sample Sizes and Data Sources**

	Lottery winners (Good taxpayers)	**Lottery non-winners (Good taxpayers)**
Sample Size and Data Sources	Admin. Data, $N \approx 7,200$* (+ Surveys, N=1,000)	Admin. Data, $N \approx 7,200*$ (+ Surveys, N=1,000)

Total N=2,000 (Survey data); $N \approx 14,400$ (Administrative data). * The municipality estimates there are approximately 7,200 winners of the lottery since 2004 but we believe the number may be smaller. Our sampling procedure will ensure that the number of lottery winners approximately equals the number of eligible non-winners.

Our approach has the advantage that the treatment and control groups will be approximately the same size (balanced design), which is typically the most efficient design conditional on the overall size of the study group.[13] Moreover, the procedure naturally distributes the study group across the four types of taxes in proportion to the prevalence of winners of lotteries for each tax. Here we in fact have a series of mini-natural experiments, in which each lottery generates a treatment group of winners and a control group of non-winners. Thus this is effectively blocked random assignment, where the blocks are individual lotteries; however, the probability of winning any lottery is the same in every block (1/10,000), so we will not need to adjust for blocked assignment in our analysis. Finally, because our treatment and control groups are both random samples from the population of eligible ("good") taxpayers in each corresponding lottery, we can use our natural experiment to estimate population average treatment effects (PATEs) for the population of good taxpayers.

This design allows us to use straightforward comparisons to estimate the effects of winning a lottery, among good taxpayers. For example, mean differences between the columns of Table 3.1 estimate the average causal effect of winning the lottery.[14] Varying the number of included tax years subsequent to

[13] Due to our sampling method, the size of the treatment and control groups are random variables; however, this will not lead to bias in treatment effect estimators due to independence of the denominator and the ratio of the numerator to the denominator, in estimators of treatment effects such as the average causal effect.

[14] We may also estimate spillover effects by comparing average payments of winners' and non-winners' neighbors. This is

the date of each lottery allows us to assess the persistence of effects. We discuss the analysis of the natural experimental data further in Section 5.

3.2 Field experiment: positive vs. negative incentives

The comparison of winners and losers likely underestimates program impact, because the lottery may induce some bad taxpayers to bring their accounts up to date to gain eligibility for the lottery. Moreover, the lottery has apparently not been effectively advertised by the municipal government, which raises the question: if Montevideo—or another municipal government—were to use an informational campaign to tell citizens about the existence of the rebate lottery, what sort of interventions would be most effective in boosting tax payments? Finally, what mechanisms may explain any effect we find of winning the lottery on tax payment?

To answer these questions, and to probe basic motivations for tax compliance, we use a field experiment in which we provide varied information to a random sample of taxpaying households. Our informational experiments allow us a unique opportunity to compare the effects of positive incentives provided by the lottery to the effects of negative incentives: for example, messages about sanctions such as fines for non-payment?

3.2.1 Text of informational treatments

In more detail, we collaborated with the municipal government to design and mail to households flyers printed with messages that correspond to the following treatment conditions:

1. *Placebo control* (reminder that tax bill is due);

2A. *Individual reward 1* (informing citizens of existence of lottery);

2B. *Individual reward 2* (also priming probability of winning);

3. *Individual sanction* (existence of fines/ punishment for non-payment);

attractive as some of the informational effect of the lottery could conceivably work through neighbor-to-neighbor communication. However, this introduces non-trivial logistical obstacles because it involves linking a large number of geo-located physical addresses to taxpayer records—rather than sampling taxpayer records that are linked to physical mailing addresses, as we do here. We are uncertain whether the inferential benefit would outweigh the cost, as we have other ways (described next) to estimate the effects of learning about the lottery's existence on future compliance. If we decide to estimate spillover effects to neighbors, we will file an amendment to this pre-analysis plan in advance of the additional data collection and analysis.

4. *Social reward* (emphasizing social rationale for lottery); and

5. *Social sanction* (emphasizing social rationale for fines/punishment).

The experimental realism of our treatments is very high: when folded for mailing, the logo of the municipality is visible, and upon delivery to households the flyers appear *identical* to municipal tax bills.[15] The experience of receiving and opening a tax bill on which the municipality prints encouragements to pay taxes would thus be very similar to the experience of receiving these flyers stamped with the municipal government's logo. Figures 2-7 show the messages on the Spanish-language flyers, while Figure 8 shows the back side of the flyer with the municipal logo.

The translated text of the flyers is as follows:

PLACEBO CONTROL (Figure 2):

Dear neighbor:

We want to remind you that **the second payment of property taxes is due in July**. If you have not received your bill, you can obtain a duplicate copy on our web site (www.montevideo.gub.uy).

For questions: Phone the tax department (1950 300)

INDIVIDUAL REWARD 1 (Figure 3):

Dear neighbor:

We want to remind you that **the second payment of property taxes is due in July**. If you have not received your bill, you can obtain a duplicate copy on our web site (www.montevideo.gub.uy).

The municipal government of Montevideo wants to reward good taxpayers. **If you pay on time, you will be automatically entered in a lottery to win a year free of property tax payments.**

Lotteries occur every other month of the year in conjunction with the National Lottery. The winners will be duly informed and the results of the lottery will be published on the web site of the city government.

You can be the next winner!

For questions: Phone the tax department (1950 300)

[15] For example, our survey interviewers exclaimed that the folder flyers appeared to them to be tax bills.

INDIVIDUAL REWARD 2 (Figure 4):

Dear neighbor:

We want to remind you that **the second payment of property taxes is due in July**. If you have not received your bill, you can obtain a duplicate copy on our web site (www.montevideo.gub.uy).

The municipal government of Montevideo wants to reward good taxpayers. **If you pay on time, you will participate automatically in a lottery to win a year free of property tax payments.**

In each lottery, 1 of every 5,000 households receives this benefit.

Lotteries occur every other month of the year in conjunction with the National Lottery. The winners will be duly informed and the results of the lottery will be published on the web site of the city government.

You can be the next winner!

For questions: Phone the tax department (1950 300)

INDIVIDUAL SANCTION (Figure 5):

Dear neighbor:

We want to remind you that **the second payment of property taxes is due in July**. If you have not received your bill, you can obtain a duplicate copy on our web site (www.montevideo.gub.uy).

Those who do not pay on time may be subject to fines and charges. The municipal government of Montevideo may take legal and administrative actions to enforce the rules where applicable.

Pay on time, avoid fines and charges!

For questions: Phone the tax department (1950 300)

SOCIAL REWARD (Figure 6):

Dear neighbor:

We want to remind you that **the second payment of property taxes is due in July**. If you have not received your bill, you can obtain a duplicate copy on our web site (www.montevideo.gub.uy).

The municipal government of Montevideo wants to reward good taxpayers. **If you pay on time, you will participate automatically in a lottery to win a year free of property tax payments.**

Lotteries occur every other month of the year in conjunction with the National Lottery. The winners will be duly informed and the results of the lottery will be published on the web site of the city government.

The municipal government of Montevideo conducts this lottery to recognize good taxpayers for their contribution to constructing a city that is more just and better for all.

For questions: Phone the tax department (1950 300)

SOCIAL SANCTION (Figure 7):

Dear neighbor:

We want to remind you that **the second payment of property taxes is due in July**. If you have not received your bill, you can obtain a duplicate copy on our web site (www.montevideo.gub.uy).

Those who do not pay on time may be subject to fines and charges. The municipal government of Montevideo may take legal and administrative actions to enforce the rules where applicable.

Fines and charges are a sanction to those who do not pay their taxes and do not contribute to constructing a city that is more just and better for all.

For questions: Phone the tax department (1950 300)

The varied messages printed on our flyers allow us to study the following topics, to which we return in Section 5, Hypotheses and Tests.

Positive vs. negative incentives. How do the effects of the positive incentives provided by the lottery compare to the effects of negative incentives/sanctions? One previous informational experiment found that emphasizing fines and other legal consequences of non-payment, in a message similar to our Individual Sanction condition, increased tax compliance in Argentine municipalities by more than 4 percentage points.[16] As discussed in the next sub-section, our design replicates this intervention and compares it to the effect of information about the rebate lottery, among both good and bad taxpayers. In a developing-country context in which tax enforcement is routinely difficult, the effects of information about sanctions should provide a useful benchmark for comparing the effects of information about rewards. (For both treatments, we will use instrumental-variables analysis to study the effects on those who *learn* about sanctions and rewards through our interventions—see Section 5).

Informational effects. Winning the lottery provides a temporary income boost, and it may shape broader attitudes towards the fairness and social benefits of taxes; yet, it may also affect future tax compliance by shaping knowledge of the lottery and perceptions of the likelihood of future rebates.[17] To estimate pure informational effects, we compare good (eligible) non-winners who receive a flyer informing them of the existence of the lottery to a control group of eligible non-winners. We can then compare this to the effect of winning the lottery among eligible taxpayers, which combines informational effects with other factors that may shape compliance behavior. (Again, an instrumental-variables estimator gives an estimate of the effect of learning about the lottery for uninformed good taxpayers influenced by our informational treatment—see Section 5).

Measuring effects among bad taxpayers. A limitation of the government's program, from an inferential point of view, is that the lottery is restricted to good taxpayers. To assess the broader impact of the program—e.g., the effect of giving bad taxpayers greater incentives on the margin to pay their taxes and thus gain eligibility for the lottery—we use our informational intervention to estimate the effect of informing ineligible ("bad") taxpayers of the existence of the lottery, compared to the placebo control group (using both intent-to-treat and instrumental-variables analysis). While "bad" taxpayers may constitute a relatively small group during good economic times (say, 10-15% of taxpayers, though

[16]Carlos Scartascini and Lucio Castro. 2013. Tax Compliance and Enforcement in the Pampas: Evidence from a Field Experiment. Manuscript. Abstract at http://www.cscartascini.org/work-in-progress.

[17]Winning lotteries may also engender self-reinforcing beliefs about individual merit, which might have broader political implications. See e.g. Di Tella, Rafael, Sebastian Galiani, and Ernesto Schargrodsky, 2007, The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters, Quarterly Journal of Economics, 122: 209-41.

we will have more precise estimates later), during crises the size of this group can grow substantially; moreover, assessing affects for bad taxpayers may give some insight into likely effects of similar interventions in settings with larger numbers of bad taxpayers than Montevideo.

Our data analysis also helps assess other hypotheses, as discussed in Section 5.

3.3 Treatment assignment

To create our study group for the informational experiment, we asked the government to draw a random sample of 28,500 administrative records, including 14,250 taxpayers who are eligible to win the lottery for exoneration of payment starting in July 2014 ("good taxpayers"), and 14,250 who are ineligible based on not being up to date on payments over the past year ("bad taxpayers").[18] We then randomized these eligible and ineligible taxpayers with equal probability to one of the six treatment groups. The sample sizes shown in Table 3.1.[19]

Due to cost considerations, we will only survey a random sub-sample of households in each treatment group, rather than a census of the 28,500 households; thus, in the Placebo Control group among good taxpayers, we sent flyers to 2,850 households, but will gather household survey data for a random sample of 1,000 of these households. We will assess effects of our treatments on tax compliance using administrative data on all 28,500 households, while we assess effects on attitudes, knowledge of the program, and other outcomes using data from our household survey.

There are two important wrinkles. First, the list of addresses provided by the municipality includes non-physical persons (e.g., corporations) as well as physical persons living in households. The municipality does not have a ready way to distinguish these persons (which is why they must screen lottery-winning taxpayer accounts before awarding a year free of tax payments). Thus, our survey firm will also screen these addresses, interviewing only physical heads of household and/or physical persons who have responsibility for paying household taxes. We initially gave our survey firm a list of 8,000 randomly selected households, which it will use to fill the quota of 6,000 households in the field exper-

[18] We worked with the government on the technical requirements of drawing the sample; when we have the requisite data, we will compare covariates of our random sample of taxpayers to covariates in the population of taxpayer records, adjusting for the oversampling of ineligibles.

[19] Note that we sometimes conceptualize treatments 2A and 2B as the same treatment (i.e., we often pool across these treatment conditions), so the overall size of the 2A plus 2B group is kept equal to the other four groups.

Table 3.2: **Field Experiment: Treatment Conditions and Sample Sizes**

	Lottery non-winners (Good taxpayers)	**Sample of ineligibles (Bad taxpayers)**
1. Placebo Control	Admin. Data, N=2,850 (+ Surveys, N=1,000)	Admin. Data, N=2,850 (+ Surveys, N=1,000)
2A. Individual Reward/ Lottery	Admin. Data, N=1,425) (+ Surveys, N=500)	Admin. Data, N=1,425 (+ Surveys, N=500)
2B.(+ Probability of Winning)	Admin. Data, N=1,425) (+ Surveys, N=500)	Admin. Data, N=1,425 (+ Surveys, N=500)
3. Individual Sanction	Admin. Data, N=2,850 (+ Surveys, N=1000)	Admin. Data, N=2,850 (+ Surveys, N=1000)
4. Social Reward*	N=2,850 (Admin. Only)	N=2,850 (Admin. Only)
5. Social Sanction*	N=2,850 (Admin. Only)	N=2,850 (Admin. Only)
TOTAL N	Admin. Data, N=14,250 (+ Surveys, N=3,000)	Admin. Data, N=14,250 (+ Surveys, N=3,000)

Total N=6,000 (Survey data); N=28,500 (Administrative data). * For these conditions, only administrative outcome data will be gathered. See sub-section (3.2.1) for text of flyers in the information experiment.

iment, after eliminating non-physical persons and accounting for non-response; should this list not be sufficient to reach the desired sample size, the firm will return to us for additional randomly-selected households. We will follow the same procedure for household surveys in the natural experiment. Control over the process of random substitutions from the list of 28,500 taxpayer addresses will allow us to calculate the real rate of non-response as well as to estimate the proportion of physical persons among the population of taxpayer accounts.

Second, we learned only after obtaining contact data (household addresses) for our 28,500 sampled households and mailing our flyers to them in late June 2014 that an important number of taxpayers in our sample (6,789 accounts, or around 25%) pay their taxes in entirety for the full year, or for other reasons do not pay taxes in July.[20] Data analysis shows these are predominantly good taxpayers—and moreover they cannot be readily influenced by the treatments in our field experiment because they will not pay taxes in our initial period of outcome measurement. We therefore plan to exclude these households from both our household survey and administrative data analysis. Thus, the population

[20]For example, pensioners pay taxes in November.

from which our household survey and administrative data samples are drawn should be conceptualized as “all taxpaying households with bills due in July 2014.”[21] Note that these are to a greater extent “taxpayers at risk” who may be affected by our intervention, in that the excluded group that does not pay in July consists disproportionately of good taxpayers.

One important issue here is a loss of statistical power relative to our planned design (see section 3.4), given that some of the households in our original sample households do not pay bills in July and thus cannot be affected by our intervention (in the short run, at least). We therefore amended our protocol to collect another round of administrative data to fill out our 28,500 cases with additional accounts paying taxes in July. We then randomized these cases to treatment groups and mailed flyers in two additional sub-phases, a Phase II.A with flyers mailed beginning July 5, 2014 and arriving by July 10, and a Phase II.B mailed from July 11 to July 26 (see sub-section 3.3.1). Because our flyers must arrive in advance of the tax due date, in this Phase II we screened records to include only those taxpayers who receive their tax bills after July 10, 2014. Thus, for Phase II the population from which we sampled taxpayer records should be conceptualized as “all taxpaying households with bills due after July 10, 2014.” Assignment of due date is not randomized; however, qualitative evidence suggests it is haphazard. For example, the zones used by the municipality to order mailings of tax bills do not correspond to postal or other jurisdictional demarcations, and the mailing does not appear to follow a strong geographic logic. We will use geo-coding of the zones to map the order in which tax bills are sent, and compare covariates of individuals who receive tax bills earlier and later; we can also compare estimates from Phase I data (in particular, households with bills due prior to July 10 and those due later) to see if these estimates are statistically distinguishable. We think that obtaining internally valid estimates for a larger group of taxpayers is a first-order concern and this prompts our collection of these additional Phase II data.

3.3.1 Timing of intervention and data collection

Figure 1 describes the timing of our field experimental intervention, the municipality’s mailing of tax bills, and data collection. Our flyers/informational treatments were generated, addressed, and then distributed by a company we hired beginning on June 27 and continuing until July 1 (Phase I). On

[21] From this population, we draw a stratified random sample with an oversample of ineligible/bad taxpayers.

July 7, we received from the postal service a list of addresses where these flyers were not received (so we could replace these addresses for purposes of the household survey). As discussed in the previous sub-section, we began the Phase II (A and B) mailings of flyers on July 5, targeting households with intermediate tax due dates for Phase II.A and the latest tax due dates for Phase II.B.

We can be confident that the households in our study group received our informational treatments before the due date for paying taxes—and in most cases, before the physical receipt of tax bills. Mailing of tax bills by the municipality began on July 1 and continues until July 21 in a staggered fashion, according to which different zones of the city receive the tax bill at different times. The due date for tax payments is also staggered and follows the order of delivery of tax bills, with the due data approximately 8 days after the receipt of bills. We were not initially aware of how the staggering would occur, and the city zones used by the municipality do not correspond to postal codes; thus, we did not time the mailing of our informational treatments to coincide exactly with the arrival of the tax bill. However, all households in the Phase I mailings of informational treatments should have received flyers before July 8, the earliest possible due date. Moreover, for Phases II.A and II.B, we gathered information on due dates for each household from the municipality and grouped households into Phases II.A and II.B accordingly, to ensure that flyers would arrive in advance of due dates and (usually) tax bills. The greater risk is that in some cases (especially Phase I mailings) our flyers could arrive several days or even weeks in advance of tax bills and due dates; this lag between receipt of our informational treatments and receipt of tax bills and due date—and our even later measurement of attitudinal outcomes using household surveys—could weaken the effect of our treatment, for households where the lag is long. However, we will have data on the date at which bills were received and the due date for payments; so in principal we can compare households who received informational treatments closer to the due date with those where the lag was greater. [22]

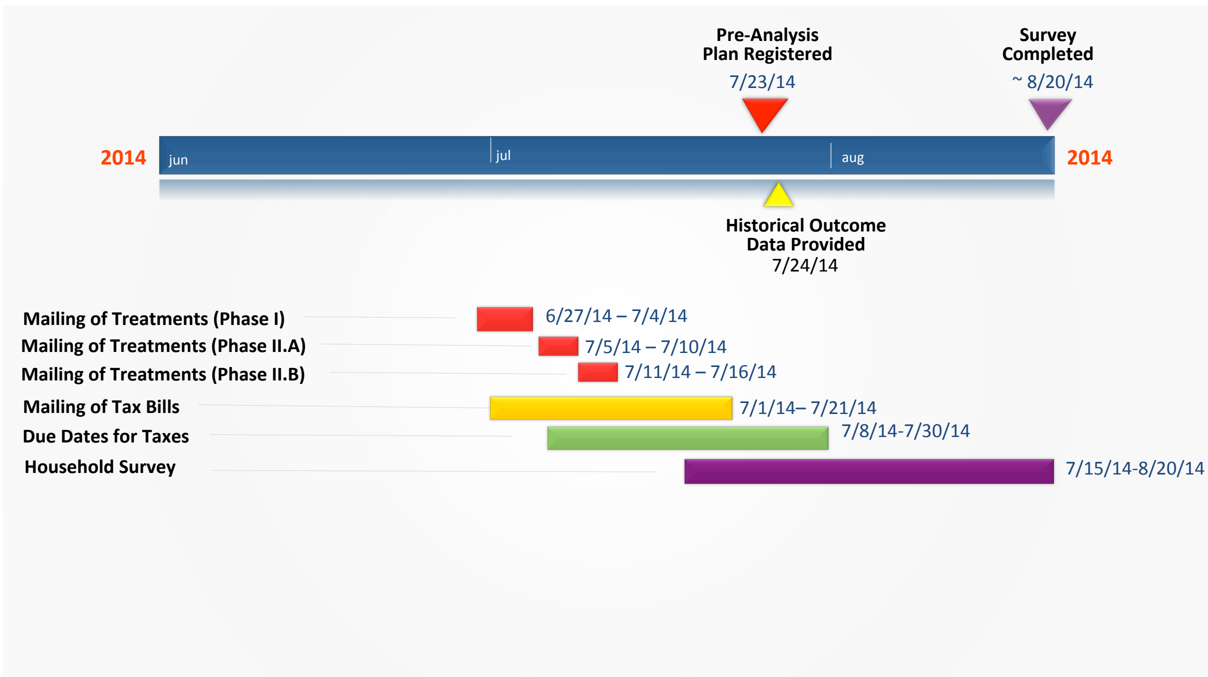
Historical outcome data for the natural experiment will be received from the municipality on approximately July 24. Our survey was fielded beginning on July 15 and will continue through approximately August 20, at some point after which we will receive survey outcome data from CIFRA. This pre-analysis plan is registered effective July 23, 2014.

Thus, we are filing this pre-analysis plan after the beginning of our field experimental intervention

[22] See discussion of analysis of heterogeneous effects in Section 5.

and household survey, but before any outcome data are collated or analyzed.

Figure 1: Timing of intervention and data collection



3.4 Power calculations

In this sub-section and the associated Figures 9-12, we present formal justification for the sample sizes in our natural and field experiments (Tables 3.1 and 3.2). For several of our power analyses, we take as a benchmark the informational experiment of Castro and Scartascini (2013), who estimate effects of informational treatments on tax compliance of over 4 percentage points using *negative* incentives for compliance (reminding taxpayers of fines for non-compliance, as in our Individual Sanction treatment).[23] However, we calculate the probability of rejecting the null hypothesis of no effect, given various true effect sizes. This effect could be, e.g., the difference in subsequent tax compliance rates for lottery winners and eligible non-winners, or differences in various graded outcomes measured through administrative or survey data.

Binary outcomes (e.g. tax compliance): There are N units with n_T units assigned to treatment and $n_C = N - n_T$ to control; we assume equal numbers assigned to treatment and control ($n_C = n_T$), as in our natural experiment per Table 3.1 and as in most of of the pairwise comparisons in Table 3.2 for the field experiment. We suppose average tax compliance is around 70%, thus the variance of this binary outcome is $0.7 \times (1 - 0.7)$, pooling across treatment and control groups. Thus, the standard error for the difference of tax compliance rates across treatment and control groups is[24]

$$\sqrt{\frac{0.7 \times 0.3}{n_T} + \frac{0.7 \times 0.3}{n_C}}, \tag{5}$$

or, using $n_T = n_C = \frac{N}{2}$,

$$\text{SE} = \frac{2\sqrt{0.7 \times 0.3}}{\sqrt{N}}. \tag{6}$$

For each effect size, we calculate power under a two-tailed test as

$$1 - \Phi(2 - \frac{\textbf{effect}}{\text{SE}}), \tag{7}$$

where Φ is the normal cumulative distribution function, SE is given by equation (6), and **effect** is the true effect size.[25] Equation (7) gives the approximate area above the normal curve centered over **effect** that is more than two standard errors away from 0, the effect size under the null hypothesis.[26] For a one-tailed test, we use 1.65 in place of 2 in equation 7; a one-tailed test is more appropriate for many of our unidirectional hypotheses discussed in Section 5 (e.g., knowledge of the lottery increases

[23]Lucio Castro and Carlos Scartascini. 2013. "Tax Compliance and Enforcement in the Pampas: Evidence from a Field Experiment." Manuscript. Abstract at http://www.cscartascini.org/work-in-progress.

[24]We use the "conservative" formula for the standard error in randomized experiments, which is the same as for the difference of proportions of two independent samples; for formal justification, see Appendix notes 31, 33 of David Freedman, Roger Pisani, and Roger Purves, 2007, *Statistics*, W.W. Norton Co., 4th edition.

[25]We switch the signs in (7) to give the area *greater* than two standard errors above zero.

[26]To be conservative, here we use 2 in place of 1.96, though we can rely on the central limit theorems and use normal approximations for most hypothesis tests; with smaller n, one might want to use the t-distribution or permutation tests.

tax compliance, among lottery losers or bad taxpayers, but does not decrease it).

Figures 9 and 10 show power for two-tailed and one-tailed tests, respectively, assuming true effects of 4, 6, 8, and 10 percentage points, e.g., **effect** $\in \{0.04, 0.06, 0.08, 0.10\}$. In each figure, the vertical line shows the study group size, pooled across treatment and control groups, that is needed for 80% power given each effect size. For $N = 2,000$, we have slightly more than 80% power given a true effect size of 6 percentage points, using a two-tailed test; for a one-tailed test, we have 80% power against a true effect size of 5 percentage points ($N = 2,000$). With a one-tailed test, we also have 80% power for an effect size of 6 percentage points when $N = 1,500$. These calculations suggest reasonable power to measure moderate effects with binary outcomes, using our survey data. However, to measure the binary outcome of tax compliance, we will use cheaper administrative data and thus a larger N, so our power will be substantially greater.

Graded outcomes: Power is greater with graded measures rather than binary outcomes. Our household survey will measure attitudes towards the tax system, often using scales instead of binary outcomes (e.g., degree of agreement with statements about the fairness of the tax system); and we also construct graded measures of indebtedness, as discussed in Section 4. In Figures 11 and 12, we measure effect sizes in relation to the unknown standard deviation of this outcome variable. Thus, for pooled $N = 2,000$, we will have power of just over 80% against a true effect size of 0.13 standard deviations (two-tailed test).

Estimating control-group parameters: One important role of the household survey is to allow us to estimate the proportion of taxpayers in Montevideo who are uninformed about the existence of the lottery. It is critical that we estimate this proportion precisely, as this estimated proportion is the denominator in some of our instrumental-variables analyses. We have several sources of data to estimate the proportion of uninformed taxpayers: (1) survey data on the placebo control group in our field experiment ($N = 2,000$, pooling across eligibles and ineligibles); (2) survey data on the "individual sanction" group in our field experiment, who we also do not inform of the existence of the lottery ($N = 2,000$, pooling across eligibles and ineligibles); and (3) survey data on the control group in the natural experiment ($N = 1,000$).[27]. If 50% of good taxpayers who have never won the lottery

[27] The latter group is a random sample of the population of eligible taxpayers, but eligibility is as of the date of different lotteries, so this is a stratified random sample where we do not know the probability of being in each strata; we may therefore opt to use only (1) and (2) to estimate the proportion of uninformed taxpayers.

are unaware of its existence, the standard error for our estimate of this population percentage is 0.79% pooling across good and bad taxpayers and using data from (1) and (2) ($N = 4,000$). (We will need to weight the estimates when pooling, to account for our oversampling of ineligible taxpayers). With $N = 5,000$, the standard error is 0.71%.

Justification for sample size. These power calculations justify our sample size for the household surveys, as depicted in Tables 3.2 and 3.1. Our power is about 80% against effect sizes for tax compliance comparable to those estimated in previous research, in the case of negative incentives. Our sample size gives us similar power against movements of around 0.15 standard deviations in attitudinal dependent variables measured as scales. Finally, our sample of households who have not won the lottery allows us to estimate the proportion of taxpayers who are uninformed about the lottery with fairly good precision; these estimates are important for assessing overall program impact as well as the likely effects of more effectively promoting knowledge of the tax rebate lottery.

With respect to the administrative data collection, our sample size balances our desire for more data against the cost in time and effort to the municipality.[28] One issue is that some of the tax payer records are for juridical not physical persons (i.e., they are companies). In our surveys, we have the ability to filter juridical persons, ultimately by visiting households; this will allow us to assess the overall proportion of physical persons in the population of taxpayers. However, this will certainly diminish our true power, relative to these calculations. We therefore want to err as much as possible on the side of a large sample size for the administrative data.

4 Outcome Measures

4.1 Administrative (tax payment) data

The policy we study originated in 2004, creating a rich time series of tax payment data for lottery winners and non-winners. We have permission from the municipal government to access historical records for our sampled taxpayers from 2000 to 2014. In addition, we will estimate treatment effects in our field experiment using data on tax compliance posterior to the mailing of our informational

[28]Supplying data is not costless for the municipality, as will involve manual extraction of records from municipal databases using the four-digit IDs we generate, and the municipality will only grant access to a sample of the data.

treatments. Our study period will continue until at least December 2015 to allow us to assess the persistence of effects of our informational interventions using administrative records.

The administrative records allow us to define three main measures of tax compliance as outcome measures:[29]

1. *Compliance (0-1)*: This is a dichotomous indicator for whether a given taxpayer account is fully paid as of the due date. We measure this outcome at each payment date, i.e., three times a year for property and vehicle tax and six times a year for the sewage and head tax.

2. *Missed Payments*: This variable measures the total number of missed payments as of a given tax due date. It varies among taxpayers who are not fully paid up and can increase or decrease during each tax payment period.

3. *Total Debt*: This variable measures the total amount of debt, including unpaid principal, interest, and charges, at each tax payment due date.[30]

In addition, we will collaborate with the municipal government to obtain an additional outcome measure. Our informational treatments (including the placebo control) prompt tax payers to log on to the municipality's website to obtain duplicate copies of their tax bills if needed. We will use the web log and records of the municipality to define the following measure:

1. *Web bill request*: This measures whether a taxpayer logged on to view his or her tax account, change information such as the mailing address associated with the account, or print a duplicate copy of the bill.

The municipality will provide us only with the date or dates at which account information was accessed in a given time period (not the content of the activity).[31] Accessing the web site can be viewed as a measure of *intended tax compliance*, or at least interest in tax records, and it is therefore an interesting outcome variable for a number of our informational treatments.

[29]We may be able to define additional outcome measures prior to analysis, in which case we will amend the pre-analysis plan.
[30]Due to complications involved in calculating historical debts (e.g., the interest and charges that applied at a date in the past), this measure will be only available for our field experiment and only in the period immediately following our intervention.
[31]Web access requires entering a tax account number; thus, the municipality can track date of access by account number.

4.2 Survey data

Our survey instrument gathers data on individual covariates and the main attitudinal dependent variables in our analysis: for example, perceived equity and fairness of the tax system; perceptions of the benefits provided by taxation; and a host of attitudinal (e.g., belief in individual merit) and political (e.g. support for the incumbent Frente Amplio) variables that may be affected by winning the lottery or by knowledge of the lottery. These data on beliefs and perceptions will help us assess mechanisms that may explain any effects we estimate, as discussed in the proposal. The survey instrument will also measure knowledge of the rebate lottery, whether respondents know anyone who has won the lottery, and related variables, which will allow us to assess the likely effects of advertising the tax rebate program more widely.

Here we register eight main outcomes:

1. Trust in municipal government (question C.4.1 in our survey instruments);[32]
2. Trust in civil servants (question C.4.3);
3. Evaluation of the mayor (question C.4.4) (and/or Performance in Office of Mayor, C.9);
4. Fairness of municipal taxes in general (question D. 6); and
5. Fairness of the property tax (question D.7.1).
6. Attitudes towards tax amnesties (question D.8)
7. Agreement that sometimes taxes are not worth paying (question D.9)
8. Party vote intention (question E.5), or change in vote intention from previous election.

Most of these survey questions are all measured on a 0-10 scale. In our instrument, there are a host of other outcomes related to these items as well as to the performance of government and quality of public services. We do not register those secondary outcomes but will certainly explore them descriptively.

4.2.1 Survey experiments

We also will evaluate several outcomes in connection with our survey experiments. First, for the survey experiment about fines and charges, we register the following outcomes, where respondents are asked

[32]Note that the order of questions varies across versions of the questionnaire, due to our survey experiments, but the question identifier/number does not.

for their degree of agreement on a 0-10 scale with the following statements:

1. "People only pay their taxes on time when there are substantial fines and charges" (survey question M.1.1);
2. "In Montevideo, punishments don't apply to the privileged" (question M.1.4); and
3. "Fines and charges for bad taxpayers are pointless" (question M.1.5).[33]

Next, for the survey experiment about the benefit of tax holidays, we register these outcomes (again, as survey questions that ask for degree of agreement on a 0-10 scale):

1. "Policies that reward good taxpayers are a waste of money" (question S.1.1);
2. "In Montevideo, benefits for good taxpayers always go to the same people (question S.1.4)."[34]

Finally, we also asked outcome questions that are identical for both the "fines and charges" treatments and the "benefit of tax holidays" treatments. As for other outcome measures, we will compare the effects of variation in the treatments to assess effects. However, these questions, which are repeated in the corresponding sections of the survey instrument, will also allow us to use the survey experiment to compare the effects of perceptions of negative vs. positive incentives directly:

1. "In general, the municipal government does a good job" (questions M.1.3 and S.1.2);
2. "In Montevideo, it is worth it to be up to date on ones taxes" (question M.1.2 and S.1.3)'[35]
3. "How would you classify the taxes that the municipal government charges, in general: very just, fairly just, a little just, or not just at all?" (questions M.1.6 and S.1.5)[36]

5 Hypotheses and Tests

In this section, we discuss general hypotheses derived from our theoretical discussion and describe our operationalization of these hypotheses, measurement of key outcome variables, and key statistical tests. For a number of our hypotheses, we also describe "mechanisms"—which we understand here as

[33] "*Las multas y recargos a malos paradores no sirven para nada.*"
[34] "*En Montevideo, los beneficios para buenos pagadores se los llevan los mismos de siempre*."
[35] "*En Montevideo, vale la pena estar al día con los impuestos.*"
[36] "*Cómo clasificara los impuestos que cobra la Intendencia de Montevideo en general: muy justos, bastante justos, poco justos o nada justos*?"

intermediate outcomes that could be shaped by our natural, field, and survey experimental interventions and that may help explain any broader program impacts we identify. The impact of our treatments on these intermediate outcomes is therefore of interest for shedding light on mechanisms. (Note that we do not intend to do formal mediation analysis, as the assumptions needed for mediation methods to identify causal effects are too demanding in our context (as in many).[37] However, we use variation-in-treatments in our field experiment design to shed light on reasons any effects of winning the tax holiday lottery, as well as the broader impact of the lottery policy).

Tables 7.3 and 7.4 list the data source and outcomes that we use to test each hypothesis, as well as the specific operationalization of each test; we discuss each hypothesis and test in detail next.

5.1 Impact of the tax holiday lottery

Hypothesis 1A: Winning the tax holiday lottery leads to an *increase* in future tax compliance.

Hypothesis 1B: Winning the lottery leads to a *decrease* in future tax compliance.

Hypothesis 1C: Winning the lottery leads to *no change* in future tax compliance.

> **Operationalization of 1A, 1B, and 1C–natural experiment**: Comparison of compliance and missed payments after the tax holiday lottery takes place. We conduct this analysis over the entire period for which data are available after time t at which each respective lottery is held (thus, outcomes are measured at times $t + \ldots$, as indicated in the first row of Table 7.3). We also compare total debt of winners and eligible non-winners as of the end of July 2014 (the only date at which we measure total debt for reasons discussed previously).
>
> **Test 1: K-S test** We construct a plot in which the horizontal axis is time, measured in tax holiday lotteries (which occur every two or four months, depending on the tax). The vertical axis is one of our three measures of tax payment (compliance, missed payments, or total debt). The date at which each taxpayer in the study group won the tax holiday lottery (treatment group) or was eligible to win that lottery (control group) is centered at 0 on the horizontal axis. We then us a K-S test for the equality of distributions. The treatment and control groups should be balanced before 0, due to the randomization provided by the

[37] On mediation, see e.g. Gerber and Green 2012.

lottery. Differences between the distributions after 0 indicates a treatment effect.

Test 2: Diff-in-diff. We calculate the average value of payment at $t + x$ in the treatment and control groups, and subtract the average value at $t - 1$ (or $t - 0$) for each group. Then, we compare the change in average outcomes in the treatment (winners) and control (eligible non-winners) groups; this is a difference-in-differences analysis. Standard errors are calculated using the conservative variance formula for the difference-in-difference.[38]

Test 3. Persistence effects. We vary x in the calculation above to estimate the persistence of effects, and compare effects for x less than the average value observed in our data set to x greater than the average value.

There are two mechanisms associated with Hypothesis 1A, a positive effect of winning a tax holiday.

Mechanism 1A.1: *Informational*. The municipal government appears to advertise the existence of the lottery quite poorly, and winning the lottery provides information about its existence (as well as a year free of tax payments). Thus, winning the lottery provides taxpayers with information that they have a positive probability of winning in the future if they pay taxes promptly.

Operationalization 1: Field experiment. Comparison of tax compliance and web bill requests of households that receive our lottery/individual reward treatments 1 and 2 (pooled together) and the control group, separately among eligible and ineligible households. A positive effect for eligible households suggests that part of the impact of the lottery on future compliance is due to the information that winning provides. A positive effect for ineligible households suggests larger program impact, as knowledge of the lottery causes bad taxpayers to pay on time (become good taxpayers).

Test 1: Diff.-in-Diff. We compare the change in outcomes for the treatment group (whom we inform of the existence of the lottery) and the placebo control group, using two outcome measures: compliance (0-1) and whether taxpayers viewed their accounts online (web bill request).[39] For the treatment condition, here we pool the three groups that are informed

[38] See e.g. Thad Dunning, 2012, *Natural Experiments in the Social Science: A Design-Based Approach.* Cambridge University Press, Chapter 6.

[39] We will need to have pre-intervention data for the web bill request to do the diff-in-diff.; the municipality has told us it can provide historical data on web requests.

of the lottery: Individual reward 1; Individual reward 2; and Social reward. (Elsewhere, we distinguish effects for these groups). This is intent-to-treat analysis, because we do not take into account whether tax payers already know about the existence of the lottery prior to receiving our informational treatment. We measure the post-intervention outcome at $t+1$—i.e., the next possible tax payment, which occurs in July 2014—as well as $t+4$, the payment that occurs in July 2015. The reason for the latter is that we will next compare effects in the field experiment to effects in the natural experiment.[40] In the natural experiment, the first recorded outcome for lottery winners occurs at $t + 4$, since winners have a year free of tax payments and thus cannot opt to comply or not comply with taxes until $t + 4$.[41] We do this analysis for both eligible and ineligible taxpayers.

Test 2: Instrumental-variables analysis (IV). Intent-to-treat analysis estimates the effect of providing citizens with information, regardless of whether taxpayers already know about the existence of the lottery. Yet, not all non-winners are ex-ante uninformed. Thus, here we also use instrumental-variables analysis. Dividing the estimated average treatment effect of information (in the difference-in-differences analysis) by the difference in information rates of bad taxpayers in treatment and control groups provides an instrumental-variables estimate of the effect of information on "Compliers"—those who learn about the lottery from our intervention.[42] We estimate the denominator—the proportion of uninformed good taxpayers—using our survey data for good taxpayers in the field experiment's placebo control and Individual Sanction conditions, since these treatments do not inform taxpayers about the lottery."[43] This analysis is important for policy, because it identifies

[40] Note that strictly speaking, the effects are only estimated for the same population with the small natural experiment occurring in July 2014, i.e., winners compared to non-winners who are eligible to win as of that date. However, we might think of estimating effects for "eligible taxpayers" in general, including those who were eligible at earlier points in time.

[41] A subtlety here is that $t + 1$ represents the first opportunity for taxpayers in the field experiment to pay their taxes, while $t + 4$ is the first opportunity in the natural experiment; thus, in the latter, more time passes between intervention and outcome. However, we have confirmed that winners of the tax holiday receive bills every three months with a zero balance during the year that they are exonerated from paying taxes. This makes the time from intervention to outcome more equivalent, if one thinks of receiving the final notice of zero tax due as a part of the treatment.

[42] Recall from section **??** that some of our flyers were mailed to juridical persons (e.g. corporations), rather than physical persons. The denominator of the IV estimator must take this into account, i.e., the proportion of informed or uninformed households is constructed taking into account the entire study group including juridical persons.

[43] As discussed previously, we could also use the control group in the natural experiment, but that introduces some complications we mentioned.

the marginal impact among taxpayers of an informational campaign to advertise the policy more widely.[44] Again, we do this analysis for both eligible and ineligible taxpayers.

Operationalization 2: We compare the effect on future compliance of information in the field experiment—among eligible households—and the effect of winning the lottery, in the natural experiment. If effects of winning are entirely informational, these two effects should be statistically indistinguishable.

Test 1. We compare the estimated effect of the "existence of lottery" treatments in the field experiment (relative to the placebo control) to the estimated effect of winning the lottery in the natural experiment, using data at $t + 4$ in both cases, and test whether the estimated effects are significantly different.

Operationalization 3: We compare respondents whom we inform through the survey experiment about the existence of the lottery (individual and social benefit) with those not informed about lottery.[45]

Test 1. We compare means on the survey question measuring whether respondents say it is "worth it" to pay taxes.

Mechanism 1A.2: *Attitudinal.* The second mechanism works through taxpayers' faith in the equity and transparency of the tax system, which may boost their willingness to pay taxes (i.e., provide an expressive benefit b as in our decision model). As noted in the introduction, allocating public benefits (such as tax holidays) through lotteries represents a distinctive form of *programmatic politics*—one in which binding, public criteria guide the distribution of benefits (and benefits are not conditional on political support of beneficiaries).[46] This form of distributive politics is particularly egalitarian, because all eligible taxpayers have the same probability of winning benefits.

Beyond tax compliance, allocating public benefits (such as tax holidays) through lotteries represents a form of distributive politics that is quite distinct from the conditionalities of clientelism or

[44] Note that IV analysis estimates effects for a particular type: uninformed bad taxpayers who would learn about the lottery from an informational campaign. Yet, such a campaign should be targeted at exactly this population.

[45] Note that in designing the study, we did not originally conceive of the discretion treatment as a pure control group in this sense. But the information about "discretion" is fairly light (though the municipality is said to "choose" winners), so this may work.

[46] For this definition of programmatic politics, see Stokes et al. (2013).

patronage politics. For political incumbents, rewarding good taxpayers in an equitable and transparent way may also have important electoral consequences, particularly in a context in which amnesties for *bad* taxpayers are common and potentially unpopular. We treat the impacts of the lottery on attitudes towards the transparency and fairness of the tax system as potential mechanisms that may help explain any impact of the lottery on tax compliance. Yet, we are also interested in the attitudinal impact of the tax holiday policy on political perceptions, independent of any consequences for tax payment behavior. In addition, for political incumbents, rewarding good taxpayers in an equitable and transparent way may also have important electoral consequences, particularly in a context in which amnesties for delinquent taxpayers are common.

Note that these are not hypotheses about the impact of winning a *particular* lottery, nor are they hypotheses about being informed that one might win the lottery; rather, they are hypotheses about how the *existence of the policy* shapes attitudes and behavior. Thus, while we conceptualize these attitudinal effects as *mechanisms* that might explain the impact of the lottery policy on tax compliance behavior, effects of the policy on political attitudes and support for incumbents is interesting regardless of whether this in turn induces greater tax compliance. We therefore characterize the following hypothesis about the impact of the tax holiday policy on perceptions of the state and the character of public policy: the tax holiday lottery increases perceived transparency and fairness of the tax system as well as political support for incumbents responsible for the policy.

> **Operationalization:** We assess whether being informed about the lottery and/or winning the lottery boosts 1. trust in the municipality; 2. trust in civil servants;[47] 3. evaluations of the mayor; 4. perceived fairness of municipal taxes in general; 5. perceived fairness of property taxes (field experiment—for the natural experiment, we use the specific tax for which the taxpayer had eligibility); and 6. degree of agreement with amnesties for bad taxpayers. Regarding the 6, note that good taxpayers who are current on their tax payments may resent amnesties for bad taxpayers; a tax holiday lottery that rewards good taxpayers may therefore influence their perceptions of the tax system as well as of political incumbents. It may also cause them to look more favorably on tax amnesties.

[47] We first ask about the municipal union, then other municipal civil servants; we use the latter question.

Test 1. For both the field and natural experiment, we conduct difference-of-means tests comparing the treatment and control groups.

Test 2. We also test for heterogeneous effects—in the case of the field experiment, comparing eligibles to ineligibles, and in the case of the natural experiment, comparing recent to old winners.

Test 3. We also asses whether the effect of *winning* the lottery on attitudes towards the fairness of the tax system and support for incumbents will be [greater than/equal to] the effect of knowledge of the lottery for eligible taxpayers. Taxpayers who have actually won the lottery have received a concrete benefit (in the form of a year of exoneration from taxes), whereas those we inform about the existence of the lottery have merely received information about the policy. Thus, if the effects on attitudes towards the state and support for political incumbents for both groups are similar (and non-zero), it suggests that the policy mainly works by shaping attitudes about the desirability of this mode of distribution.

Regarding Hypothesis 1B, at least two mechanisms might explain a negative effect of winning the lottery on future compliance:

Mechanism 1B.1: *Income effects*. Winning the lottery gives taxpayers a year free of paying taxes; in principal, this additional income could buttress them against the costs of punishment in case of non-payment for future taxes.

Operationalization: We assess whether the effects of winning the natural experiment–which provides a year free of income–varies according to the cost of payment, here operationalized in terms of the property value.

Test: Heterogeneous effect analysis—difference of means.

Mechanism 1B.2: *Behavioral/"habit" effects*. Taxpayers who stop paying taxes for a year may fall "out of the habit," leading them to miss making tax payments once they are due.

Operationalization: We assess whether the effects of winning the lottery vary by time since winning.

Test: Natural experiment—heterogeneous effects by time since winning.

Operationalization: The behavioral effect should not exist for taxpayers who are set up to pay taxes by automatic debit (placebo outcome) **Test**: Field and natural experiment—assess effect for those who self-report using automatic debits to pay taxes.

Finally, Hypothesis 1C—a null effect of winning the lottery—could follow, for instance, if the lottery has only informational effects on future compliance, but good taxpayers are fully informed about the lottery before winning. In this case, winning the lottery would not inform them about a possible future benefit of staying current on their tax payments. Moreover:

Mechanism 1C.1: *Erroneous beliefs*. Taxpayers who have won the lottery may believe, falsely, that their probability of winning a future lottery is lowered by having won previously. This could either lead to a decrease in tax compliance among winners, relative to eligible non-winners, or to a null effect of winning the lottery. In particular, taxpayers who already knew about the existence of the lottery before winning it might comply more before winning the lottery than after doing so.[48]

Operationalization: Using a survey question, we assess whether taxpayers believe that someone who has won the lottery has less of a chance of winning it in the future.[49] Then, we compare the effects of winning according to beliefs in the dependence of winning on previous outcomes.[50]

Test 1: Heterogeneous effects of winning the lottery, by beliefs about independence of lotteries: compare those who say a winner is less likely to win

[48]That is, some taxpayers who were previously informed about the lottery may have increased compliance in order to boost the probability of winning; but having won, they may believe erroneously that their chances of winning are reduces, providing a disincentive to continued compliance.

[49]Survey question S.5: If a person wins [the tax holiday] lottery, would you say that the chances of winning again on another occasion are: greater, the same, or less?

[50]Note, however, that beliefs in the dependence of winning are post treatment.

again, to those who say the chances of winning a second lottery are equal or greater for someone who has already won.

A final possibility under H1.C is that only expressive benefits (e.g. a sense of duty) matter for payment of taxes—and winning the lottery has no effect on these perceived expressive benefits of compliance.

5.2 Rewards vs. punishments.

Hypothesis 2A: (Priming knowledge of) sanctions and punishments for non-compliance *increases* tax compliance.

Operationalization: We compare payments at $t + 1$ (July 2014) of taxpayers who are reminded/informed of fines and punishments for non-payment to the placebo control group.[51]

Test 1: Here, we pool the "individual sanction" and "social sanction" treatments, as both treatments remind taxpayers of the existence of fines and punishments for non-payment and we are using administrative outcome data which are available for both groups. (Later, we distinguish between the effects of these treatments). The test statistic is the difference of means, divided by the estimated standard error; as elsewhere, we use the conservative Neyman estimator of the variance (and standard error). Since we have a directional hypothesis (one-tailed test), we reject the null at the nominal 0.05 level if this statistic is greater than 1.65 (though see below for adjustments for multiple comparisons).

We also assess one possible mechanism:

Mechanism 2A.1: *Beliefs about probability of punishment*: the sanction treatments increase beliefs in the probability of punishment. Also, the positive effect of reminders is stronger for those who believe the probability of punishment is greater.

[51] Unfortunately, we do not have a survey question that asks individuals about the existence of fines and sanctions, which would allow us to estimate the proportion who are informed about these.

Operationalization: First, we assess whether the sanctions treatment decreases the proportion of taxpayers who say it is "very probable" or "somewhat probable" that the municipal government offers an amnesty to bad taxpayers in the coming year, relative to those who say it is "a little probable, or very improbable." Next, we hypothesize that the effects of the sanction treatment will be weaker for those who say it is very probable or somewhat probable (weak belief in the probability of punishment) to those who say it is a little probable, or very improbable (strong belief in the probability of punishment).

Test: First, we use a difference-in-means test to assess whether the sanctions treatment decreases belief in the probability of an amnesty (one-tailed test). Then, we text whether the estimated effects of the sanction treatments are higher for the group that believes the probability of an amnesty is low, using a one-tailed test. Note that beliefs in the probability of an amnesty are plausibly post-treatment for taxpayers in the treatment (sanction) group; so this is not heterogeneous effects analysis but rather a type of mediation analysis. Note that the assumptions needed for to identify this mediation analysis are strong.

We are especially interested in comparing the size of the effects of manipulating negative and positive incentives to pay taxes, though we are agnostic about the direction of the effect:

Hypothesis 2B: The absolute value of the effect of priming knowledge of sanctions and punishments is [greater/less than] the effect of priming knowledge of the tax holiday lottery.

Operationalization: Here, we are interested in the effects on tax payment at $t + 1$ (July 2014) of the rewards treatments vs. the sanctions treatments. Here, we pool the "individual reward 1," "individual reward 2" and "social reward" as the "rewards treatments" and "individual sanction" and "social sanctions" as the "sanctions treatment."

Test 1: We compare the estimated effects in a difference-in-difference analysis. The estimated standard error for the diff-in-diff is the square root of the sum of the estimated variances of the two estimated effects. We use a two-tailed test.

We also hypothesize that the effects of manipulating both positive and negative incentives to pay taxes are heterogeneous, depending on taxpayers' compliance history.

Hypothesis 2C: *Marginal taxpayers I.*The effects of both rewards and sanctions are greater for marginal taxpayers (those with only some history of non-compliance) than for those with no history of late payments or those with an extensive history of late payments.

The rationale here is that the cost of becoming a good taxpayer is greatest for those who are seriously in arrears; they are less likely to be moved by the promise of a one-year tax holiday, conditional on bringing all their accounts up to date.[52] Taxpayers with an extensive history of late payments may also be less responsive to the threat of sanctions.

On the other side of the spectrum, taxpayers with no history of late payments may always pay on time, due to intrinsic motivations (a large expressive benefit b to tax payment); for such taxpayers, paying taxes on time is part of their social identity, and their taxpaying behavior is unlikely to be shaped by the field or natural experimental interventions.

Thus, those taxpayers most likely to be affected are those on the margins: either taxpayers who are currently in arrears, but not very much, or those who are currently eligible for tax holiday lotteries, but who have been in arrears in the past.

> **Operationalization**: We operationalize the idea of marginal taxpayers by identifying "good taxpayers" with some history of past non-payment,[53] and "bad taxpayers" (current ineligibles) who do not owe more than 3 quotas of past tax payments.[54]
>
> **Test 1**: Among good taxpayers, we compare the effect of information about the existence of the lottery, and the effect of information about fines and sanctions, across those good taxpayers who have never owed a debt and those who have at some point in the past been past due on their taxes.[55] Among bad/ineligible taxpayers, we compare the effects for those with 3 or fewer quotas owed to the effects for worse taxpayers. In both cases we use one-tailed tests.

[52]We will use our sample of taxpayer records to characterize how common these very delinquent taxpayers are.

[53]We will seek to identify these taxpayers using historical payment data on the field experimental study group.

[54]We choose the threshold of three payments because the municipality allows taxpayers who owe no more than three quotas to continue to pay their current tax bill without accumulating interest or charges on the 3 quotas that are past due.

[55]As noted previously, we have screened out of our study group taxpayers who pay the full year's taxes at once and thus do not pay in July. These taxpayers tend overwhelmingly to be good taxpayers. Thus we are already looking at relatively marginal taxpayers in our study group.

The idea that marginal taxpayers will be most responsive to manipulating perceived benefits and costs suggests a final hypothesis as well:

Hypothesis 2D: *Marginal taxpayers II*. The effects of both rewards and sanctions are greater for taxpayers who face lower costs of coming into compliance.

> **Operationalization**: Taxpayers with high incomes but low property values (and therefore taxes) face lesser costs in tax payments in general and should be less responsive to manipulation of costs and benefits.[56] On the other side of the spectrum, taxpayers with low incomes and high property values and thus taxes should face greater costs of compliance and should thus also be less responsive to treatments.[57]. Those taxpayers who are on the margin—those intermediate ratios of incomes to taxes owed—should be most responsive to our field and natural experimental treatments. We therefore operationalize the cost of coming into compliance in terms of ratios of income to property values, and we hypothesize that the effects are weaker for taxpayers for whom these ratios are extreme.
>
> **Test 1.** We compare the effects on compliance for taxpayers within one standard deviation of the average value of income-to-property-value ratio[58] with the effects for other taxpayers.

5.3 Individual vs. social incentives.

Our informational interventions will allow us to compare the costs and benefits of a campaign to increase awareness of the tax holiday program, as well as the effects of reminding taxpayers of sanctions for non-payment. Yet, an important question is whether emphasizing the social benefits and costs makes such a campaign more effective. In terms of our theory sketched above, the question is whether manipulating the value of the expressive benefit b can boost the impact on tax compliance. Thus, we have the following hypotheses:

Hypothesis 3A: (*Social benefits*). Emphasizing social rationale for *benefits* provided through the tax holiday lottery increases compliance, relative to an emphasis on individual returns.

[56] An observational corollary is that these taxpayers should tend to be current on their accounts.
[57] These taxpayers should tend to be in arrears more often
[58] Alternately, those from the 25th to 75th percentile on this measure?

Operationalization: We compare the effect of the social benefit treatment to the effect of the individual benefit treatment (or alternately, simply assess whether the mean response for the social benefit treatment is greater/different than the mean response for the individual benefit treatment).

Test 1: Difference-of-means; two-tailed test (since in principle social rational could also *decrease* tax compliance, relative to the individual reward. Here, for the individual reward, we pool individual reward 1 and individual reward 2.

Hypothesis 3B: (*Social sanctions*). Emphasizing the social rationale of *sanctions* for non-payment increases compliance, relative to an emphasis on individual punishment.

Operationalization: We compare the effect of the social sanction treatment to the effect of the individual sanction treatment (or alternately, simply assess whether the mean response for the social sanction treatment is greater/different than the mean response for the individual sanction treatment).

Test 1: Difference-of-means; two-tailed test (since social rationale for sanctions could also *decrease* tax compliance, relative to a focus on individual sanctions.

If both hypotheses 3A and 3B were born out in the data, it would provide evidence that social preferences—operationalized in terms of manipulating the expressive benefit of paying taxes, or social cost of not paying taxes—play an especially important role in shaping compliance behavior. Such a finding could inform the design of informational interventions by the municipality of Montevideo or similar governments.

5.4 Adjustments for multiple comparisons

We will present both nominal p-values and corrected p-values, using a false discovery rate (FDR) correction to control the Type-1 error rate. For both our natural and field experiments, we will control the FDR at level 0.05.[59]

[59]The problem of multiple comparisons also arises with our survey experiments, as we have multiple outcomes for each experiment. We will take a similar approach to adjustment in that case.

Thus, for a given randomization with m (null) hypotheses and m associated p-values, we order the realized nominal p-values from smallest to largest, $p_{(1)}, \leq p_{(2)} \leq \ldots \leq p_{(m)}$. Let

$$k \text{ be the largest } i \text{ for which } p_{(i)} \leq \frac{i}{m} 0.05.$$

Then, we reject all $H_{(i)}$ for $i = 1, 2, \ldots, k$, where $H_{(i)}$ is the null hypothesis corresponding to $p_{(i)}$.[60]

For comparison, we will also present strict Bonferroni corrections, i.e., for each hypothesis $H_{(i)}$, we reject the null at the adjusted 0.05 level if $p_{(i)} \leq \frac{0.05}{m}$. This correction will lead to the most conservative inference for each individual pairwise comparison. Our rejection rule, however, will require controlling the overall false discovery rate.

How large is m under our study design? This differs for the field, natural, and survey experiments. For the natural experiment, we have one randomization into treatment (winner) and control (eligible non-winner) groups. (As discussed, this is really blocked randomization, where the blocking is by type of lottery; however, the blocks are all in expectation the same size for each type of tax). Meanwhile, we will have nominal p-values associated with each of the following comparisons:

- K-S test (three outcomes: compliance, missed payments, and total debt)
- Diff-in-diff (three outcomes)
- Persistence of effects, heterogeneous effects (three outcomes)
- Difference of means (six outcomes: trust in municipality, trust in civil servants, evaluation of mayor, fairness of taxes, fairness of the tax specific to the corresponding lottery,[61] and opinion of lottery)
- Heterogeneous effects, by cost of payment (three outcomes)
- Heterogeneous effects, by time since winning (three outcomes)
- Heterogeneous effects, by beliefs about non-independence of winnings (three outcomes)

[60] For a description of this procedure, see Yoav Benjamini and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological).* 57 (1): 289-300.

[61] We will do these separately for the winners/eligible losers for each type of tax, i.e. focus on the fairness-of-property-tax outcome onlyfor those who won or were eligible to win a year free of property taxes.

The total number of comparisons is 24. We also have the p-value for the comparison of effects in the natural and field experiment. This makes a total of $m = 25$ p-values for the natural experiment.

For the field experiment, we have the following comparisons:

- Diff-in-diff (two outcomes each for $t + 1$ and $t + 4$).[62]
- Difference of means (six outcomes: trust in municipality, trust in civil servants, evaluation of mayor, fairness of taxes, fairness of property taxes, and and support for amnesties—the analysis is conducted for both eligibles and ineligibles, so twelve tests total);
- Effect of punishments treatment on compliance and belief in probability of punishments (two outcomes), and comparison to effect of reward (one outcome);
- Heterogeneous effects, by payment history (one outcome)
- Effect of social benefits (one outcome), effect of social sanctions (one outcome), comparison of effects (one outcome)

The total number of comparisons is 23. We also include the p-value for the comparison of effects in the natural and field experiment. This makes a total of $m = 24$ p-values for the field experiment.

6 Relevance, Contribution, and Value of Research

Promoting tax compliance is critical in developing countries, where tax monitoring and enforcement is often weak. We believe have a valuable opportunity to use a randomized policy intervention to learn about the effects of positive as well as negative incentives for compliance, as well as the separate impact of priming social versus individual rationale for punishments/benefits. Our informational intervention will allow us to conduct cost-benefit analysis on scaling up promotion of knowledge of the rebate lottery.

[62] We do not count the $p-values$ associated with the IV analyses as nominal significance should not differ from the intent-to-treat analysis. An initial report may ignore the $t + 4$ outcomes until we have gathered those data a year later.

7 Data Availability

Data will be posted upon publication of a research report or two years after the start of our informational intervention, whichever is sooner.

Figure 2: Text of informational intervention (Spanish): Placebo control

Intendencia deMontevideo

Estimado/a vecino/a:

Queremos recordarle que en el mes de **julio vence la segunda cuota de la Contribución Inmobiliaria.** Si todavía no recibió su factura, puede obtener un duplicado en nuestro sitio web (www.montevideo.gub.uy).

Por consultas:

FONO TRIBUTOS

1950 3000

Figure 3: Text of informational intervention (Spanish): Lottery/individual reward 1

Intendencia deMontevideo

Estimado/a vecino/a:

Queremos recordarle que en el mes de **julio vence la segunda cuota de la Contribución Inmobiliaria.** Si todavía no recibió su factura, puede obtener un duplicado en nuestro sitio web (www.montevideo.gub.uy).

La Intendencia de Montevideo quiere premiar a los buenos pagadores. **Si usted paga en fecha participará automáticamente de un sorteo por la exoneración de un año de Contribución Inmobiliaria.**

Los sorteos se realizan todos los meses pares del año junto con la Lotería Nacional. Los beneficiados serán debidamente informados y se publicarán los resultados en el sitio web de la Intendencia.

¡Usted puede ser el próximo!

Por consultas:

FONO TRIBUTOS
1950 3000

Figure 4: Text of informational intervention (Spanish): Lottery/individual reward 2

Intendencia deMontevideo

Estimado/a vecino/a:

Queremos recordarle que en el mes de **julio vence la segunda cuota de la Contribución Inmobiliaria.** Si todavía no recibió su factura, puede obtener un duplicado en nuestro sitio web (www.montevideo.gub.uy).

La Intendencia de Montevideo quiere premiar a los buenos pagadores. **Si usted paga en fecha participará automáticamente de un sorteo por la exoneración de un año de Contribución Inmobiliaria.**

En cada sorteo, 1 de cada 5.000 hogares recibe este beneficio.

Los sorteos se realizan todos los meses pares del año junto con la Lotería Nacional. Los beneficiados serán debidamente informados y se publicarán los resultados en el sitio web de la Intendencia.

¡Usted puede ser el próximo!

Por consultas:

FONO TRIBUTOS
1950 3000

Figure 5: Text of informational intervention (Spanish): Individual sanction

Intendencia deMontevideo

Estimado/a vecino/a:

Queremos recordarle que en el mes de **julio vence la segunda cuota de la Contribución Inmobiliaria.** Si todavía no recibió su factura, puede obtener un duplicado en nuestro sitio web (www.montevideo.gub.uy).

Quienes no paguen en fecha podrían estar sujetos a multas y recargos. La Intendencia de Montevideo podría tomar acciones administrativas y legales para hacer cumplir la normativa en los casos que correspondan.

¡Pague en fecha, evite multas y recargos!

Por consultas:

FONO TRIBUTOS
1950 3000

Figure 6: Text of informational intervention (Spanish): Lottery/Social reward

Intendencia deMontevideo

Estimado/a vecino/a:

Queremos recordarle que en el mes de **julio vence la segunda cuota de la Contribución Inmobiliaria.** Si todavía no recibió su factura, puede obtener un duplicado en nuestro sitio web (www.montevideo.gub.uy).

La Intendencia de Montevideo quiere premiar a los buenos pagadores. **Si usted paga en fecha participará automáticamente de un sorteo por la exoneración de un año de Contribución Inmobiliaria.**

Los sorteos se realizan todos los meses pares del año junto con la Lotería Nacional. Los beneficiados serán debidamente informados y se publicarán los resultados en el sitio web de la Intendencia.

La Intendencia de Montevideo realiza este sorteo para reconocer a los buenos pagadores por su contribución a la construcción de una ciudad más justa y mejor para todos/as.

Por consultas:

FONO TRIBUTOS

1950 3000

Figure 7: Text of informational intervention (Spanish): Social punishment

Intendencia deMontevideo

Estimado/a vecino/a:

Queremos recordarle que en el mes de **julio vence la segunda cuota de la Contribución Inmobiliaria.** Si todavía no recibió su factura, puede obtener un duplicado en nuestro sitio web (www.montevideo.gub.uy).

Quienes no paguen en fecha podrían estar sujetos a multas y recargos. La Intendencia de Montevideo podría tomar acciones administrativas y legales para hacer cumplir la normativa en los casos que corresponda.

Las multas y recargos son una sanción para quienes no pagan sus impuestos y no contribuyen a la construcción de una ciudad más justa y mejor para todos/as.

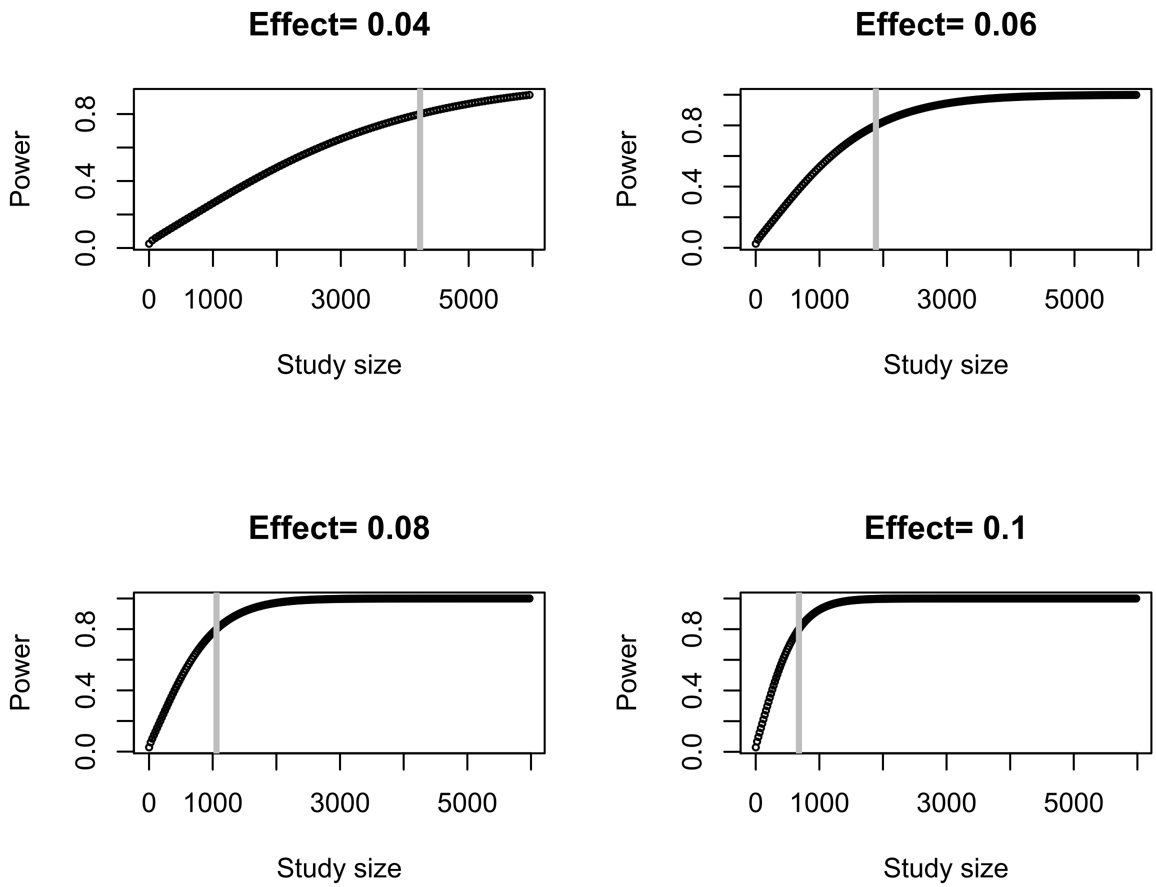
Por consultas:

FONO TRIBUTOS
1950 3000

Figure 8: Informational intervention: Reverse side of flyers with municipal logo

Figure 9

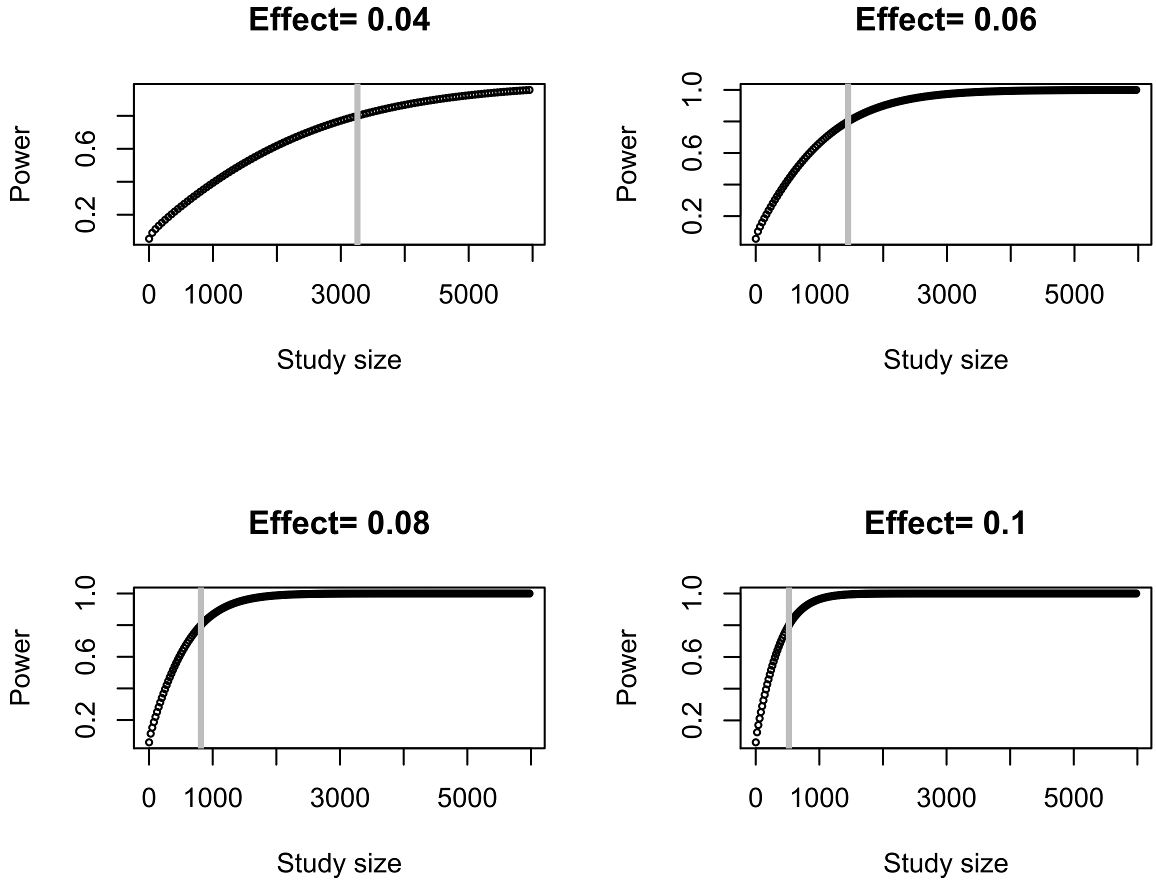
Power Calculations for Different True Effect Sizes
(Two-tailed tests, Binary Outcome)



Plots show statistical power as a function of study size for different effect sizes (binary outcome, e.g tax compliance). Effects are differences of proportions. Vertical line shows the size required for 80% power.

Figure 10

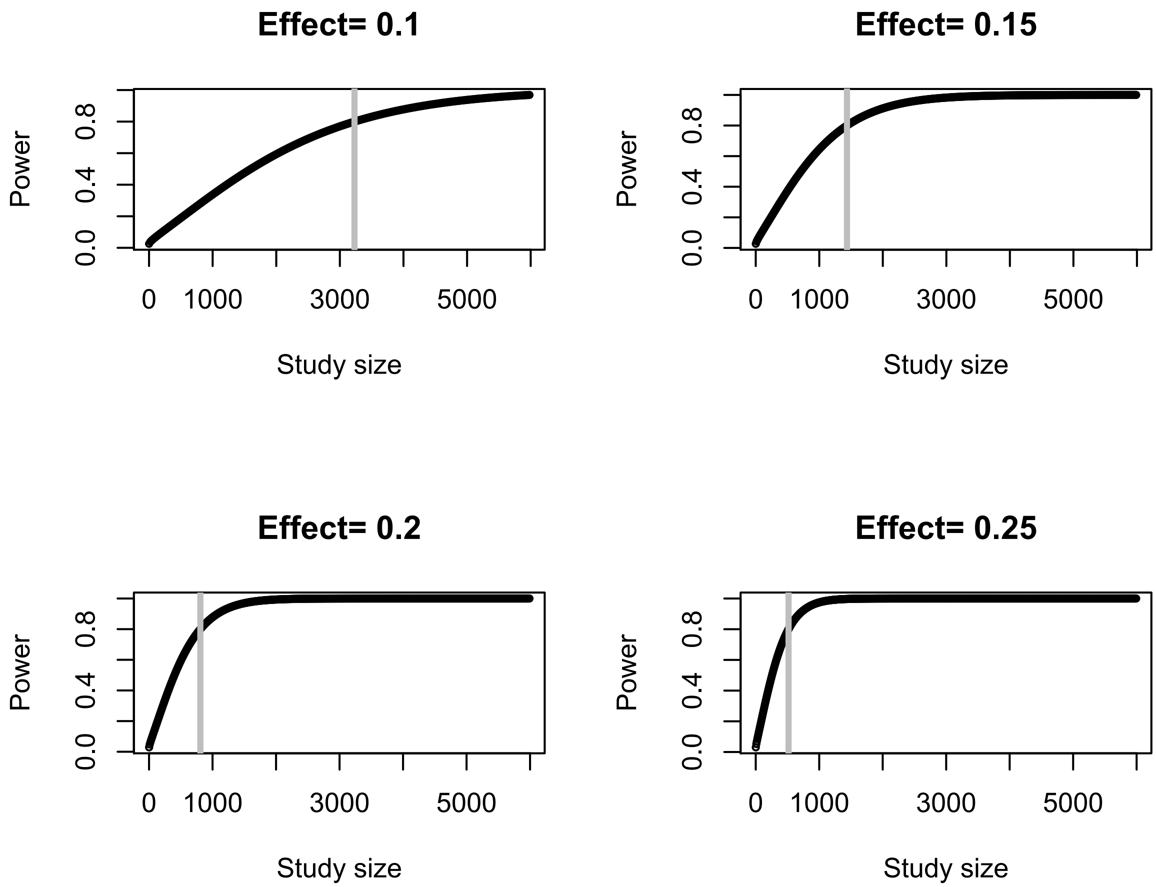
Power Calculations for Different True Effect Sizes
(One-tailed tests, Binary Outcome)



Plots show statistical power as a function of study size for different effect sizes (binary outcome, e.g tax compliance). Effects are differences of proportions. Vertical line shows the size required for 80% power.

Figure 11

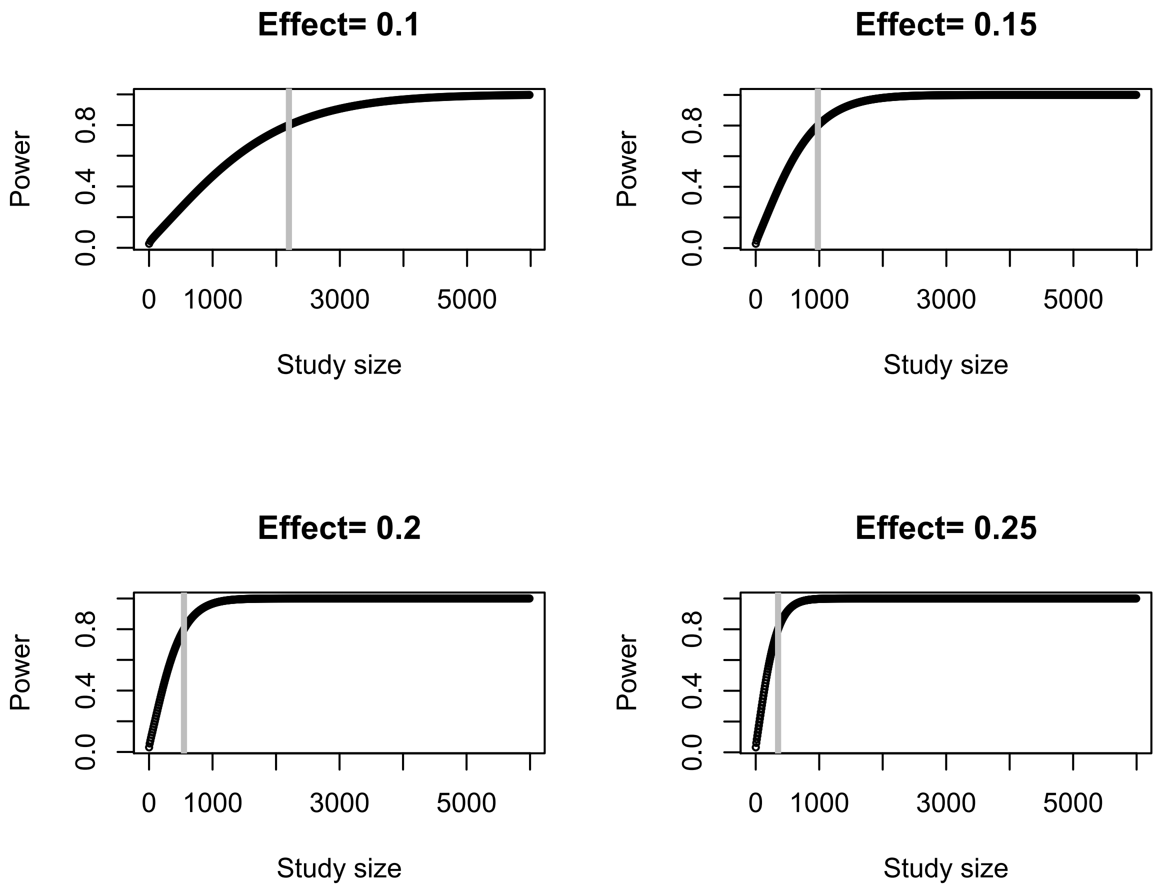
Power Calculations for Different True Effect Sizes
(Two-tailed tests, Graded Outcome)



Circles show statistical power as a function of study size for different effect sizes (graded outcome, e.g attitude scales). Effect sizes are expressed in standard deviations, e.g. 0.1 of one SD. Vertical line shows the size required for 80% power.

Figure 12

Power Calculations for Different True Effect Sizes
(One-tailed tests, Graded Outcome)



Circles show statistical power as a function of study size for different effect sizes (graded outcome, e.g attitude scales). Effect sizes are expressed in standard deviations, e.g. 0.1 of one SD. Vertical line shows the size required for 80% power.

Table 7.3: **Hypotheses, Outcomes, and Tests**

Hypotheses	Data Sources	Outcomes	Comparisons	Tests
Hypotheses 1A, 1B, 1C (Winning lottery)	Natural Exp. (Admin. data)	$t+\ldots$ 1. Compliance (0-1) 2. Missed Payments 3. Total Debt	Winners vs. Non-Winners (Eligibles)	1. K-S test 2. Diff-in-Diff 3. Persistence of effects
Mechanism 1A.1 (*Informational*)	Field Exp. (Admin data) (Eligibles and Ineligibles)	$t+1$ & $t+4$: 1. Compliance (0-1) 2. Web bill request	Existence of Lottery vs. Control	1. Diff.-in-Diff 2. IV
	Field Exp. vs. Nat. Exp. (Admin data)	$t+4$: 1. Compliance (0-1) 2. Web bill request	Effect of Info. vs. Effect of Winning	1. Diff. of Diff.-in-Diffs
	Survey exp. (Survey data)	1. Worth it to pay (Q. S.1.3)	Ex. of Lottery vs. Discretion	1. Diff. of Means
Mechanism 1A.2 (*Attitudinal*)	Field Exp (Survey data) (Eligibles + Ineligibles)	$t+1$ 1. Trust in municipality 2. Trust in civil servants 3. Eval. of Mayor 4. Fairness Taxes 5. Fairness Prop. Taxes 6. Support Amnesties	Existence of Lottery vs. Control	1. Diff. of Means 2 and 3. Diff. of Means (Het. effects)
	Nat. Exp (Survey data)	$t+\ldots$ 1. Trust in municipality 2. Trust in civil servants 3. Eval. of Mayor 4. Fairness Taxes 5. Fairness Spec. Tax 6. Opinion of lottery (Version 1 of survey)	Winners vs. Losers	1.Diff. of Means 2. Diff. of Means Het. effects – recent vs. old winners)
	Survey Exp (Survey data)	t 1. Lotteries are waste of money 3. Eval. of City Hall 4. Benefits go to "same as always"	Lottery treatments vs. non-lottery (discretion) treatment	1.Diff. of Means 2. Diff. of Means Het. effects –
Mechanism 1B.1 (*Income effects*)	Nat. Exp. (Admin. Data)	$t+\ldots$ 1. Compliance 2. Missed Payments 3. Total Debt	Winners vs. Losers	1. Heter. effects by cost of payment (property value)

Table 7.4: **Hypotheses, Outcomes, and Tests (Cont.)**

Mechanism 1B.2. (*Habit effects*)	Nat. Exp. (Admin. Data)	$t + \ldots$ 1. Compliance 2. Missed Payments 3. Total Debt	Winners vs. Losers	1. Het. effects by time since winning
Mechanism 1C.1. (*Erroneous beliefs*)	Nat. Exp. (Admin/Survey Data)	$t + \ldots$ 1. Compliance (0-1) 2. Missed Payments 3. Total Debt	Winners vs. Losers	1. Het. effects by beliefs about non-indep. of winning
Hypothesis 2A (*Punishments*)	Field Exp. (Admin. Data) Survey Exp (Survey data)	$t + 1$ 1. Compliance (0-1) t 1. Worth it to pay 2. Eval. of City Hall 3. The privileged escape fines	Existence of Fines vs. Control Fines treatment vs. lotteries treatment	1. Diff. of Means 1.Diff. of Means
Mechanism 2A.1 (*Prob. punishment*)	Field Exp. (Survey Data)	$t + 1$ 1. Belief in Prob. of Fine	Existence of Fines vs. vs. Control	1. Diff. of Means
Hypothesis 2B (*Rewards vs. Punishments*)	Field Exp. (Admin. Data)	$t + 1$ 1. Compliance (0-1)	Existence of Fines vs. Control	1. Diff.-in- Diff
Hypothesis 2C (*Marginal taxpayers I)*	Field Exp (Admin. Data)	$t + 1$	Benefit vs. Sanction	1. Het. effects by payment history
Hypothesis 2D (*Marginal taxpayers II*)	Field Exp. (Admin./Survey)	$t + 1$ 1. Compliance (0-1)	Benefit vs. Sanction	1. Het. effects by payment cost
Hypothesis 3A (*Social benefits*)	Field Exp. (Admin. Data)	$t + 1$	Social Ben. vs. Indiv. Ben.	1. Diff. of Means
Hypothesis 3B (*Social sanction*)	Field Exp. (Admin. Data)	$t + 1$	Social Sanc. vs. Indiv. Sanc.	1. Diff. of Means
Hypothesis 3A-B (*Social benefits vs. social sanctions*	Field Exp. (Admin. Data)	$t + 1$	Effect of vs. Social Sanc. vs. effect of Indiv. Sanc.	1. Diff. of Diff. of Means

In the table, t refers to tax payment periods, of which there are 3-6 per year, depending on the tax. Thus, for winners of the lottery, $t = 0$ is the period in which they won the lottery; $t = 1$ is the following tax payment period; and because they win a year free of tax payments $t + 4$ is the next payment period in which they owe taxes. Property taxes are paid three times per year.