Wp

WORKING PAPERS

DON'T BLAME THE MESSENGER. A FIELD EXPERIMENT ON DELIVERY METHODS FOR INCREASING TAX COMPLIANCE
November 2015
N° 2015/09

DON'T BLAME THE MESSENGER. A FIELD EXPERIMENT ON DELIVERY METHODS FOR INCREASING TAX COMPLIANCE

Ortega, Daniel
Scartascini, Carlos

CAF DEVELOPMENT BANK OF LATIN AMERICA

Wp

WORKING PAPERS

DEVELOPMENT BANK OF LATIN AMERICA

DON'T BLAME THE MESSENGER. A FIELD EXPERIMENT ON DELIVERY METHODS FOR INCREASING TAX COMPLIANCE

Ortega, Daniel
Scartascini, Carlos
CAF – Working paper N° 2015/09
November 2015

ABSTRACT

The effect of different delivery mechanisms for increasing tax compliance has not been evaluated so far. This study conducts a field experiment in Colombia that varies the way the National Tax Agency contacts taxpayers on payments due for income, value added, and wealth taxes. More than 20,000 were randomly assigned to a control or one of three delivery mechanisms. Results indicate large and highly significant effects, as well as sizable differences across delivery methods. A personal visit by an inspector is more effective than a letter or an email, conditional on delivery; which has several relevant academic and policy implications.

Don't Blame the Messenger. A Field Experiment on Delivery Methods for Increasing Tax Compliance

DANIEL ORTEGA AND CARLOS SCARTASCINI*

Abstract

The effect of different delivery mechanisms for increasing tax compliance has not been evaluated so far. This study conducts a field experiment in Colombia that varies the way the National Tax Agency contacts taxpayers on payments due for income, value added, and wealth taxes. More than 20,000 were randomly assigned to a control or one of three delivery mechanisms. Results indicate large and highly significant effects, as well as sizable differences across delivery methods. A personal visit by an inspector is more effective than a letter or an email, conditional on delivery; which has several relevant academic and policy implications.

* Ortega: Development Bank of Latin America (CAF) and IESA. (email: dortega@caf.com) Scartascini: Inter-American Development Bank. (email: carlossc@iadb.org). We would like to thank the participants in the 3rd TARC workshop at University of Exeter, the CIAT Tax Studies and Research Network Meeting in Montevideo, and seminars at DIAN, the Inter-American Development Bank in Bogotá and Washington, DC, and King's College London for their comments and suggestions, and to Martín Ardanaz, Raquel Bernal, Matías Busso, Paul Carrillo, Phil Keefer, Christos Kotsogiannis, Giulia Mascagni, Pablo Sanguinetti, and Christian Traxler for very fruitful discussions about this project. We would also like to thank the staff at DIAN and the Government of Colombia for their courage and collaboration, especially Maria Isabel Palomino, without whom this project would not have been possible, and Lesbia Maris, María Franco Chuaire, Edgar Castro, and Mónica Mogollón for their research assistance. We thank the Public Capacity Building Fund (KPC), funded by the Government of Korea for providing partial funding for this project. The opinions presented herein are those of the authors and do not represent the official position of their institutions. The policy evaluated in this document was executed under the norms and regulations of the Government of Colombia, and according to the procedures imposed by its National Tax Agency in its own code of ethics.

I. Introduction

Why do people pay taxes? What are the best mechanisms for collecting outstanding tax obligations? The empirical literature has advanced steadily in the last few years in trying to explain what motivates individuals to pay their taxes in full and on time, and what is the best way to deal with those who do not declare the full tax amount or are late with their payments. In particular, there has been a recent increase in the number of studies that rely on sending different types of messages to the taxpayers to identify which type of messages and content elicit a higher behavioral response from the taxpayer (e.g., Blumenthal, Christian and Slemrod, 2001; Castro and Scartascini, 2015; Chirico et al., 2015; Del Carpio, 2014; Dwenger et al., 2014; Fellner, Sausgruber and Traxler, 2013; Hallsworth, List and Metcalfe, 2014; Kleven et al., 2011; Ortega and Sanguinetti, 2013; Slemrod, Blumenthal and Christian, 2001).[1]

Almost every paper to date has used physical letters as the method of communicating the message. In this paper, we innovate and evaluate instead the effect of providing the same message but using different delivery methods, which could have a sizable impact on compliance.[2] There are many reasons for this. First, "actions may speak louder than words." Taxpayers understand that the tax agency has a menu of options that include cheap impersonal alternatives that it could use to reach the universe of taxpayers, and costlier, more personal visits that can only be used to reach a subset of taxpayers. The type of method the agency uses to inform taxpayers about their outstanding liabilities and warn them about the consequences of not paying serve as a signal to the taxpayer regarding the

[1] The tax evasion literature is extremely vast to be summarized in this paper. For comprehensive overviews, of the theoretical literature see Traxler (2010), and Hashimzade, Myles and Tran-Nam (2012). Dell'Anno (2009) and Luttmer and Singhal (2014) review the literature on the moral determinants of compliance. Hallsworth (2014) and Mascagni (2014) present broad overviews of the use of field and laboratory experiments for increasing tax compliance

[2] Evaluating different delivery methods has been common in the "get out the vote" literature. IDB (2011) surveys that literature as well as the use of new information technologies on individual behavior in several other policy areas, such as banking, and health. Haynes et al. (2013) show the effect of text messages on the payments of delinquent fines. Kessler and Zhang (2014) summarize the differential effects of methods in the health literature.

probability of being effectively forced to pay. Second, a personal visit may generate different behavior than the more impersonal methods because of social forces that make people behave differently when confronted with other people. Third, the likelihood of delivering the message effectively may differ by method.

For testing the effect of delivery mechanisms, we conducted a field experiment in Colombia in which taxpayers received a message about their due tax payments (declared but unpaid taxes).[3] Around 21,000 taxpayers who had not paid their taxes on time (commonly referred as tax delinquencies) were randomly assigned to one of three different treatments (physical letter, e-mail, inspector visit), and to a control group. Because of issues of one-sided noncompliance with the assignment to the treatment (for example, some people did not receive the messages because their address was incorrect or because the agency could not get to them within the frame of the exercise), we estimate both ITT and TOT/LATE effects.

The results in the paper confirm results already in the literature showing that sending deterrence messages has large effects on eliciting payments, and those results highlight that the method of contacting the taxpayer is relevant for explaining compliance: differences across methods could be tenfold. Personalized visits are more effective than emails, and these work better than traditional mail (and much better of course than doing nothing), conditional on delivery. Complementing these results with those in Ortega and Scartascini (2015), which shows that phone calls have an effect that would fall in between the effect of personal visits and the email we estimate here, we can conclude that personalized methods outperform the impersonal methods, which is consistent with the analytical framework pushed forward in this paper.

[3] Making people pay their declared taxes is not only an issue relevant for developing countries. In 2006, according to an estimate by the United States Treasury Department, Americans failed to pay about $110 billion, or around 25 percent of the estimate of the total amount underpaid in that year (Pérez-Truglia and Troiano, 2015).

Among those assigned to a letter (ITT results) the probability of making a payment is 4 percentage points higher than doing nothing (control group). Given that the underlying probability for the control group is about 5 percent, sending a letter almost doubles the probability that the taxpayer would cancel part of the debt. Sending an email and scheduling a personal visit has an even larger impact (about 14 percentage points higher than doing nothing, three times higher impact.) Among those who were actually treated (TOT results) payment of outstanding debt was much higher: about 8 percent higher than the baseline scenario for those receiving a letter, 17 percentage points for those receiving an email, and about 88 points for those receiving a personal visit. That is, almost every person who received a visit by a tax inspector made some kind of payment. Overall, the economic relevance of the exercise was highly significant. The Agency recovered about 3 times more previously unpaid liabilities from the people they attempted to contact than from the people in the control group. These differences are almost ten times larger for those in the group of personal visits (attempted visits). Moreover, we find large spillover effects, with those treated making payments of other arrears too.

Of course, while reaching a taxpayer with an inspector has an impact about 10 times higher than sending a letter, the relative difference in marginal costs is higher (about 16 times). Still, in the case of Colombia, because the absolute cost is relatively low the net benefit favors the personal visit over the impersonal methods, conditional on actual delivery of treatment.

These results provide information to tax agencies that may help them choose the delivery method that could maximize recovering the most revenue at the lowest cost. However, as we discuss later, the long-term optimal warning strategy depends on taxpayers' perception of how likely it is that the warning may turn into effective enforcement. Therefore, because there is a relatively fixed amount of taxpayers that the tax agency can take to the courts, making inspector

visits a universal policy may reduce their effectiveness because taxpayers may now infer that the probability of effective enforcement has dropped. On the other hand, the personal methods may increase their effect if they are used sparingly and this strategy is effectively communicated to the taxpayer who receives the warning.[4]

These results, with different effects by delivery method, are consistent with those in the donations and volunteer literature, and the "get-out-the-vote" literature (GOTV), which finds that personal canvassing and personal visits by candidates are usually more effective for getting people to vote than the more impersonal methods (Green and Gerber, 2008). This literature has substantially affected the way political parties and governments engage with their citizens. It has also stimulated the development of new analytical models for explaining voter turnout and opened up the door for new conceptualizations of how voters choose policy options.[5] We hope this paper has a similar effect for shaping tax agencies' strategies and academic research.

This paper contributes in several ways.[6] First, we show that increasing compliance and reducing delinquencies takes more than sending a persuasive message. The way the message is delivered matters too, and personal contact with the tax authority seems to be very important in the decision of whether to pay owed taxes or not. Results also show that among the impersonal methods, the email seems to be a stronger method than physical letters (about twice as important), even when accounting for the many messages that could not be

[4] For a fixed number of people that could be sent to court, a higher number of warnings reduces the probability that a "warned person" would be taken to court. Sending inspectors to every house would reduce the effect. Sending emails to a selected group and informing about it (e.g., "you are one out of 50 people we are warning") may increase the effect of the message.

[5] Barton, Castillo and Petrie (2014) have provided factual support for explaining candidates' strategy of investing heavily on personal interactions. On the theory side, Rogers, Gerber and Fox (2012) develop a conceptual model of voting as a "dynamic social expression" that integrates the results coming from the field experiments of the GOTV literature.

[6] The results in this paper are relevant primarily for the tax compliance literature but extend to the GOTV and political campaigning literature, and other literatures that evaluate the value of direct marketing, such as the growing literature on charitable fundraising (DellaVigna, List and Malmendier, 2009; Landry et al., 2006; among others), and on financial markets (Bertrand et al., 2010).

delivered. Therefore, the paper opens up the discussion in the tax compliance literature about the relevance of the delivery mechanism for affecting compliance, which may be worth including explicitly in the theoretical models. It may also make it a prerequisite of future fieldwork to be explicit about the delivery method chosen and the implications it has for experimental design (e.g., power of the experiment), and the external validity of each intervention.

Second, we show that contacting taxpayers and warning them about their outstanding debt has important spillover effects. Those in the treatment group had a higher probability of canceling the tax required by the authority and canceling other obligations too. The direction and size of the spillovers has usually been not evaluated by the extant literature.

Third, it reinforces the idea that the economic literature should consider more explicitly how policies are informed as the delivery mechanism of those policies may be as important as the content of the policies themselves for affecting individual behavior.

Fourth, the paper has relevant policy implications. First, it highlights how relevant it is for tax agencies to evaluate the way they contact their taxpayers and the potential long-term effect of each strategy, and incorporate this understanding into the cost-benefit analysis. Second, differences between ITT and TOT results stress the relevance of getting the basic things right first: having accurate, valid, and up-to-date ways to contact taxpayers may be as important in the longer run as developing other, more sophisticated enforcement strategies.[7] The cost in lost revenues may be substantial. For example, in the case of this experiment, the Agency may have had recovered an additional US$8 million approximately if they had been able to contact all the taxpayers in the treatment group (plus

[7] There are some more sophisticated enforcement strategies, which include the obligation of using electronic billing—the tax agency then is able to monitor instantaneously every transaction—and having access to third-party data such as credit card statements and the like (Corbacho, Fretes Cibils and Lora, 2013). While some countries are using some of these tools, for most developing countries there is still plenty of work ahead like implementing some of the recommendations in this paper.

additional revenues because of spillover effects). Finally, it provides evidence to governments regarding the value of communication and how different mechanism may work differently according to the policy objectives at hand.

The paper is organized as follows. Section 2 presents a summary of the related literature, and Section 3 describes the analytical framework. Section 4 describes the experiment, and Section 5 presents the empirical results. Section 6 concludes.

II. Why Might the Delivery Method Matter?

As mentioned, most of the field experiments that have tried to affect compliance through the use of messages have relied on the use of letters as the main delivery mechanism.[8] While evaluating systematically the role of different delivery technologies has been absent from the tax compliance literature, it has been more common in related literatures, such as in the *Get-out-the-vote* (GOTV) literature. Existing randomized experiments have provided relevant information on the effect of campaigning and voter mobilization on election outcomes. It has been shown that impersonal methods of voter turnout communication such as robotic calls (Green and Karlan, 2006; Ramírez, 2005; Shaw et al., 2012) and emails (Nickerson, 2006b; Stollwerk, 2006) are recurrently ineffective.[9] On the other hand, non-partisan face-to-face canvassing (Gerber and Green, 2000), and phone calls (Imai, 2005; Arceneaux, 2007; Nickerson, 2006a; and Arceneaux and Nickerson, 2006) are more effective than non-personalized methods such as flyers. This result is also confirmed by Barton, Castillo and Petrie (2014), who look at the role of candidate door-to-door canvassing. In the experiment, voters are persuaded by personal contact (the delivery method), but no evidence was

[8] In Castro and Scartascini (2015) the message was printed on the property tax bills instead of sending a letter. This method, however, would have very similar properties to sending a letter in the context of the framework we present here.

[9] Still, there is some evidence that text messages can also be effective tools to mobilize voters on Election Day (Dale and Strauss, 2009). It remains to be study the reasons behind the different effects between impersonal methods.

found for the content of the message. An emerging result from this literature, quite relevant for the research we pursue here, is that the content of the message may not be as relevant as the type and quality of its delivery for nudging people.

In this paper, we keep the content of the messages constant and evaluate the effect of different delivery methods on tax compliance. By doing so, we highlight the relevance of an issue that has been largely ignored in the literature. It also helps to compute the cost of not keeping accurate and up-to-date information about taxpayers.

One reason why the methods may have different impact is because "actions may speak louder than words." Taxpayers understand that the tax agency has a menu of options that include cheaper and more comprehensive alternatives to the personal visits. If the agency decides to visit the taxpayer to inform her of outstanding liabilities and warn her about the consequences of not paying, the taxpayer may update the probability of being prosecuted if she does not comply more than if she receives an email—which she may assume was less selective and reached more taxpayers—because: i) given a set of fixed resources, the probability of further legal action after a warning may increase with the selectivity of the delivery method; and ii) being chosen under a more selective method may indicate targeting of resources to collect her specific debts.[10] This argument can be embedded in the traditional tax evasion model (á la Allingham-Sandmo, 1972; Yitzhaki, 1974).[11]

Consider an individual taxpayer decision in a single-period setting who maximizes the expected utility from disposable income by choosing whether and how much of the debt he or she owes, T^o, to cancel, T^c.[12] The agent has a level of

10 There is another mechanism we do not exploit here fully which is that people who receive the visit of the inspector may worry that, in addition to collecting the money owed, it may lead to further inspections on the amounts declared.

11Hashimzade, Myles and Tran-Nam (2012) and Traxler (2010) constitute broad and comprehensive surveys of this literature.

12 Of course, the model could be extended in several ways, by making enforcement decisions endogenous and by evaluating both evasion and payment decisions in a dynamic setting. We could also include interaction effects across

income *Y*, and an initial amount of outstanding liabilities T^o. If the taxpayer is prosecuted because she failed to pay in spite of being prosecuted, which occurs with probability *p*, she has to pay a penalty *f* over the outstanding amount (T^o-T^c). On the other hand, if the taxpayer is not prosecuted, which occurs with probability (*1-p*) she can enjoy financial gains at a rate *r*.[13] This rate is the opportunity cost of handing over the money to the government (e.g., the interest rate gained in the market or the interest cost avoided by not having to borrow for paying the government). Consequently, *r* is individual specific.[14] Under the standard assumptions, the usual optimal decision rule equates the ratio of marginal utilities under enforcement and non-enforcement to $r_i(1-p)/fp$, which is the relative price of income in those states. Comparative statics are standard: the amount of debt canceled would be increasing in the probability of enforcement (*p*) and the fine (*f*), and decreasing in the opportunity cost of paying (*r*).

How do different delivery mechanisms affect the taxpayer decision? Assuming that prosecution can only take place after the taxpayer has been warned by the tax authority, which is the case in the context of our field experiment and in most countries, and that the tax authority has a fixed budget, *B*, that can be used either for warning actions, B^W, or for prosecuting those taxpayers who do not comply in spite of being warned about their debt level, B^E: $B = B^W + B^E$

The actual number of prosecutions, *E*, and warnings, *W*, depend on the cost of each action, *C*, and the budget dedicated to it. Therefore:

$$E = \frac{B^E}{C^E}; W = \frac{B^W}{C^W} = \frac{B - B^E}{C^W}$$

taxpayers. Still, those extensions are not necessary, given the institutional set-up in which our field experiment takes place, for the taxpayer decision we are trying to capture.

[13] The maximization problem can be written as: $Max_{Tc}\ V = pU(Y-f(T^o-T^c)) + [1-p]U(Y+r(T^o-T^c))$

[14] For example, a taxpayer who collects VAT from its customers can use this money as working capital; on the other hand, a taxpayer with low levels of income but high wealth (e.g., somebody who inherited a house) may avoid borrowing against his illiquid asset to pay the government what he owes in wealth taxes.

Then, if taxpayers can only be prosecuted once they have been warned, the enforcement/prosecution probability conditional on being warned can be written as:

$$p = \frac{E}{W} = \frac{B^E}{B - B^E} \frac{C^W}{C^E} \quad \text{and} \quad 0 \leq p \leq 1$$

Therefore, for a given budget allocated to enforcement, for those individuals contacted by the tax agency, the probability of prosecution is increasing in the cost of the warnings, C^W: $\frac{\partial p^E}{\partial C^W} > 0$

Consequently, because the taxpayer knows that personalized methods are costlier than the impersonal methods, the taxpayer will internalize a higher probability of prosecution when she receives the visit of an inspector than when she receives a letter or an email. Henceforth, those receiving the more personalized (and costly) methods should tend to be more likely to comply with the tax authority.

We could extend the model by making the tax authority's decision endogenous and by letting the taxpayer update his priors according to more complicated schemes. However, this simple framework captures the problem faced by both tax authorities and taxpayers. First, tax authorities usually have fixed bureaucratic structures and budgets, and low mobility of resources. The people who send warnings (revenue officials) are usually part of a different bureaucratic structure than the lawyers who prosecute the taxpayers. Moreover, actual prosecution usually depends on the resources assigned by courts too. Consequently, it seems very plausible that total budgets are given and the only instrument of choice by the agencies is about composition (i.e., how resources are allocated across methods). Second, taxpayers have little information about how many taxpayers are being contacted and what method the tax agency is using for

contacting other taxpayers. Therefore, assuming that taxpayers infer enforcement probabilities from the cost of each method seems a fair assumption.

A complementary reason for finding differences across methods is that receiving the visit of a tax inspector may generate different behavior than the more impersonal methods because of social forces that make people behave differently when confronted with other people. Individuals try to take actions that make others view them more favorably (Harbaugh, 1998; Lacetera and Macis, 2010), and individuals will be more likely to take action when asked to do so by someone else (Kessler and Zhang, 2014). For example, there is evidence that people are more likely to donate and volunteer when called, visited, or asked by a friend (Card, DellaVigna and Mamendier, 2011; Freeman 1997; Meer and Rosen, 2011), and more likely to vote under personal canvassing than under more impersonal methods (Imai, 2005).

Finally, there is a mechanical reason. Each method might have different probabilities of actually reaching the taxpayer and delivering the message for several reasons. The first is data quality. Not every entry in the taxpayers' record may have been updated at the same time, which can generate a different probability for reaching the taxpayer electronically or physically. A second consideration is human effort. While electronic methods are quite impersonal, physical and personal methods require the effort and dedication of mail carriers and public employees. Therefore, the effectiveness of the intervention may depend on how much human effort each treatment requires, and whether the appropriate incentives are in place.[15] A third issue to consider is taxpayer attention. Some methods require different levels of attention by the taxpayers. While a personal visit may be very salient for the taxpayer, a letter or an email may go unnoticed even if received.

[15] Even though part of the problem can be corrected in the estimations, as we show later, the researcher still depends on accurate reporting. Moreover, beyond the academic implications, reporting obviously affects policy effectiveness.

III. The Experiment

With the objective of increasing tax collection and evaluating the effectiveness of different delivery mechanisms for sending messages to taxpayers, the National Tax Agency of Colombia (DIAN) agreed to randomly assign the method used to contact a sample of taxpayers with due liabilities during one of their National Revenue Collection Days.[16]

In the context of this project, the Agency randomized a subset of taxpayers with due tax payments into four main groups. One group was assigned to be contacted via e-mail, another one via physical letter, and another group was assigned to receive the visit of an inspector. The fourth group was left as a control group.

The population of this experiment includes all taxpayers with unpaid liabilities from their income, wealth, or sales taxes for the years 2011 to 2013.[17] Taxpayers with relatively low (lower than COP20,000—about US$20 in PPP) and high (more than COP50 million—about US$46,000 in PPP) debts were not included.[18] Those who did not have a physical address, telephone number, or email on file were also left out.[19] At this point 20,818 taxpayers remained eligible. Among them, 5,000 taxpayers were assigned to standard mail, 5,000 taxpayers to email, and 4,042 to a personal visit; the remaining 6,776 taxpayers were assigned

[16] The Agency has traditionally dedicated one day every few months to trying to recoup unpaid taxes. DIAN has fewer inspectors per inhabitant than any other country in the region, which makes it harder for them to conduct massive enforcement campaigns. Running the experiment in this context had the value added of increasing the capacity of the Agency and improving the way it runs the campaigns.

[17] As Hallsworth (2014) identifies, focusing on the payment decision of a predetermined amount reduces many of the measurement problems that the papers focused on declaration have. See also Castro and Scartascini (2015) for a discussion of this point.

[18] To convert from COP to US$ in PPP terms, we use World Development Indicators' data for exchange rate (about COP1800 per dollar during the period) and PPP conversion factor (about 0.6). Data available at: http://data.worldbank.org/indicator

[19] Originally, we planned to use phone calls as an additional delivery method. Unfortunately, it could not be accomplished in the context of this experiment. Ortega and Scartascini (2015) summarizes the results of a posterior experiment which used only phone calls as delivery method.

to the control group. The randomization was performed in six blocks according to the size of debt and whether the debt was recent or not.[20]

As shown in Table A1 of Appendix, the main variables of interest are balanced across treatments using the pre-experimental data. That is, treatment groups were balanced according to the number of unpaid obligations and the amount of standing debt with the tax authority, T^o, which is the information provided to the taxpayer to affect the taxpayer's choice variable, T^c (the taxpayers decide whether to pay the informed amount of outstanding debt, a fraction of it, or nothing).

There are a few imbalances for some of treatments for some of the individual's characteristics such as being a firm or an individual—which is expected given the large number of covariates—so we include them as controls in the empirical analysis below and show that their inclusion does not affect the size or significance of the coefficients of interest. Moreover, when we interact those variables with the treatments to check for heterogeneous effects, the differences do not reflect relevant differences in behavior, as the analysis on the type of taxpayer shows (i.e., firms and individuals do not present statistically different payment behaviors after treatment).

The experiment was implemented between September and October of 2013.[21] The message included in both the physical letter and the email was exactly the same. The message stated the account balance on 31 July 2013, the type of tax, and the year or month it had not been paid. It also included information on methods of payment and the cost that the taxpayer was incurring by not paying (interest and penalties, potential legal action, and possible effect on credit history). Finally, it provided a moral suasion message ("Colombia, a commitment

[20] This way we can balance on variables that may proxy well the taxpayer relevance, economic activity, and payment history. This strategy is similar to Dwenger et al. (2014).

[21] Personal visits were carried out on 10 September 2013, emails were sent on 2 October 2013, and physical letters were sent out between 30 September and 4 October 2013.

we can't evade"). The message concluded with the contact information of a tax agency authority.[22] This way, even though the content of the messages was not the subject of the evaluation, careful steps were taken to include all the components that have been identified in the literature to matter for increasing compliance (BIT, 2012; Hallsworth, List and Metcalfe, 2014).

Personal visits had a unique protocol that inspectors were supposed to follow. At the time of the visit, if the taxpayer was present at the physical address, the inspectors identified themselves and proceeded with the protocol (included in Online Appendix). It basically followed the same logic as the written messages: the taxpayer was informed about his or her standing tax delinquencies and urged to pay. Inspectors were supposed to mention the penalties the taxpayer was incurring and the possibility of further legal actions in case of noncompliance. The visit was closed by the verbal delivery of a moral suasion message.

In the case the taxpayer was not present at the address but there was certainty that the address was correct, the inspectors left a citation informing that the inspectors had been there. In this case, no detailed information (such as the amount of debt) was left in the citation because of privacy concerns so the taxpayer was asked to visit the Tax Agency offices instead to obtain information regarding his or her standing liabilities. If the taxpayer was not present at the domicile and there was no certainty that the address was correct, then no notification was left behind. We collected the information about payments realized by the taxpayer at the end of the year.

As we discuss later, there were some cases of non-compliance with the assignment. First, the Agency didn't send the messages or could not locate some of the taxpayers. Second, in a very few cases, the Agency contacted some of the taxpayers with a method different than the one assigned during the randomization.

[22] The actual letter is included in the Online Appendix.

For these reasons, we estimate both ITT and TOT/LATE, and we also discuss its implications for external validity. Moreover, the fact that the Agency (or some of the inspectors) decided whether to accept the assignment or not provides us with the tools to investigate the determinants of behavior inside the Agency.

IV. Empirical Results

The general model we estimate is presented in the following equation

$$Y = \alpha + T\beta + X\gamma + B\delta + D\theta + \epsilon$$

where *T* is the vector of treatments (email, physical letter, and personal visit), *X* a vector of control variables, *B* the blocks (or strata), and *D* the district-level fixed effects.[23] We use several dependent variables to measure compliance. *Paid* is a dummy that takes value 1 if the taxpayer made any payment cancelling liabilities after the experiment. *Full payment* is a dummy that takes value 1 if the taxpayer cancelled the liabilities reported in the message in full. *Total Payment* is the amount (in logs) paid by the taxpayer after the experiment. *Payment share* is the share of liabilities canceled by the taxpayer. *Other payments* is a dummy that takes value 1 when the taxpayer made a payment.

The set of independent variables includes: *Liabilities*, which is the amount informed to the taxpayers in the messages; *Number of debts*, which is the number of tax obligations the taxpayers did not cancel on time; *Tax*, which is a set of dummy variables that indicate the type of tax the taxpayer had liabilities for (wealth, income, VAT); *Taxpayer type*, which indicates whether the taxpayer is a firm or an individual, *Pre-payments*, which is the amount of liabilities canceled by the taxpayer between the moment of the randomization and the experiment;

[23] Because the probability of being assigned to the control and treatment groups is not uniform across blocks we also estimate the models using weighted least squares (weights are the inverse of the probability of being selected to the control or treatment groups) even though the results are basically the same. Results are included in Online Appendix..

Wrong information, which takes a value 1 when the amount of debt informed to the taxpayer was different than his or her actual liabilities with the tax authority because of the prepayments; and *Overpayments*, which takes a value 1 in those cases when the taxpayer made a payment higher than his or her standing liabilities before the experiment took place.

As mentioned, we have six blocks defined according to the size and maturity of the debt, and district-level fixed effects (which corresponds to the geographic district the taxpayer belongs to and the tax agency jurisdiction she reports to).

IV.1 Effectiveness of the "National Revenue Collection Intervention"

The first analysis we perform to evaluate whether conducting the revenue collection exercise was worthwhile for the Agency. As shown in Table 1, during the campaign the Agency collected about COP1,800M from payments made by 335 out of the almost 7,000 taxpayers in the control group. Therefore, absent any effort by the agency (which we could call the zero deterrence scenario), approximately only 5 percent of the taxpayers would have had paid any part of their standing liabilities and only 2 percent would have had paid them in full.

Contrary to that scenario, the exercise had a large revenue collection effect for the Tax Agency. The amount it collected from people assigned to the treatment group (which we call from now on "overall treatment") was much higher: about COP8,800M (or around COP0,6M per taxpayer—almost two-and-a half-times higher than in the zero deterrence scenario, for about US$583). In the case of this group, 2,774 taxpayers made payments, which constitute 20 percent of the group, and 11 percent paid their debt in full. Importantly, there were large and significant spillovers, as 15 percent of the taxpayers canceled other obligations too.

When the same information is considered controlling for the fact that many of the taxpayers could not be located, the differences are even greater because the same payments are now drawn from a smaller taxpayer base. As we show in detail later, approximately half of the taxpayers could not be located (this average varies significantly by treatment, from 75 percent for the personal visits to 12 percent for the email). Therefore, out of the taxpayers who actually received the message stating the outstanding liabilities and the warning, the effect was about twice as high (about US$1,100 PPP per contacted taxpayer), as can be observed in the last row. If the Agency had had a more accurate database, it could have doubled the amount collected (more than US$8 million PPP).

Table 1. Summary Statistics of Intervention Results

	Taxpayers	Paid	Full payment	Total payments	Other payment
		Control group			
Total	6776	335	102	$ 1,793,000,000	0
Per taxpayer	6776	5%	2%	$ 264,610	
		Overall treatments			
Total	14042	2,774	1,519	$ 8,836,000,000	2,163
Per taxpayer	14042	20%	11%	$ 629,255	15%
Per contacted taxpayer	7457	37%	20%	$ 1,184,927	29%

Source: Authors' calculations

A summary of the regression results (OLS) is included in Table 2. In the Online Appendix we include the full set of regressions, including weighted OLS results (results are basically the same). Here, the treatment variables indicate assignment to the treatment (ITT estimates). The upper panel of the table shows the regressions results when we consider all the treatments pooled. The lower panel shows the regression results considering each treatment separately. Even columns show the results including the control variables. As can be observed, point estimates change little to none from one specification to the other.

As shown in the upper panel, taxpayers included in the treatment group had a positive and significantly higher probability of paying their liabilities (*paid*) compared to the taxpayers in the control group (10 percentage points higher) and a higher probability of paying the full amount (*full payment*)—8 percentage points higher. The share paid with regards to the informed debt (*payment share)* is 9 percentage points higher than the share paid by those in the control group, and people in the treatment group paid more than twice the amount than those in the control group (*total payment*). Interestingly, there are large spillover effects, as 13 percent of those in the treatment groups made payments to other liabilities they also had but that had not been part of the warning sent by the tax agency.

Table 2. ITT Results

	Dependent variable									
	Paid		*Full payment*		*Payment share*		*Total payment (logs)*		*Other payments*	
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Overall Treatment	0.109***	0.105***	0.076***	0.078***	0.092***	0.091***	1.469***	1.410***	0.136***	0.129***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.06)	(0.06)	(0.00)	(0.00)
N	20,818	20,818	20,818	20,818	20,818	20,818	20,818	20,818	20,818	20,818
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Letter	0.042***	0.039***	0.026***	0.027***	0.031**	0.031**	0.591***	0.550***	0.126***	0.120***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.08)	(0.08)	(0.01)	(0.01)
Email	0.153***	0.148***	0.110***	0.111***	0.135***	0.133***	2.042***	1.967***	0.139***	0.133***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.09)	(0.09)	(0.01)	(0.00)
Personal Visit	0.136***	0.133***	0.095***	0.099***	0.110***	0.110***	1.839***	1.792***	0.148***	0.138***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.12)	(0.12)	(0.01)	(0.01)
Pvalue of joint significance	0	0	0	0	0	0	0	0	0.02	0.04
Letter=Email	0	0	0	0	0	0	0	0	0.05	0.06
Letter=Visita	0	0	0	0	0	0	0	0	0.01	0.02
Email=Visit	0.10	0.18^	0.08	0.12^	0.07	0.09	0.14^	0.25^	0.36^	0.52^
N	20,818	20,818	20,818	20,818	20,818	20,818	20,818	20,818	20,818	20,818
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Each row shows the regression coefficients and the standard error in parenthesis corresponding to an OLS regression that includes strata and district. Standard errors are robust. *** p<0.01, ** p<0.05, * p<0.1

The top section of the table shows the results for a regression that includes the overall treatment variable. The bottom section shows the results for regressions that include each treatment individually.

Even columns include Liabilities (in log), Taxpayer type (firms), Type of tax dummmies, Pre-payments (in logs), Wrong Information, and Overpayments as additional controls.

^ indicates that Email and Personal Visit coefficients are not statistically different .

Source: Authors' calculations

IV.2 Relative Effectiveness of Each Delivery Method

While the overall program executed by the Agency was very important in terms of revenue collection (the revenue collected by taxpayer more than doubled), the respective effectiveness of the methods used to contact the taxpayer varies. As can be observed in the bottom panel of Table 2, personal visits (inspections) and emails were more effective than sending letters for the agency.

In terms of the economic significance, sending a letter generates a 55 percent larger amount paid (*total payment*) and increases the share of the amount paid with respect to liabilities by 3 percentage points when compared to the

control group (*payment share*). Sending a letter also favors higher compliance. On average, taxpayers in the group that were sent a letter are 4 percentage points more likely to make a payment than those in the control group (*paid*) and also 3 percentage points more likely to pay their debt in full (*full payment*). These taxpayers are also 12 percentage points more likely to make payments on other arrears they may have with the tax authority (*other payments*).

Sending an email has an even larger effect when compared to the control group. Those contacted by this method pay a 13 percentage points higher share of canceled liabilities, and they are 15 percentage points more likely to make any type of payment, 11 percentage points more likely to pay in full, and 13 percentage points more likely to make payments over other arrears not included in the experiment (spillover effects).

Scheduling a personal visit has a similarly large effect (as we show later, results are much higher when we condition for delivery).[24] Taxpayers contacted by this method pay a 10 percentage points higher share of canceled liabilities (*payment share*), and they are 13 percentage points more likely to make any type of payment (*paid*), 10 percentage points more likely to pay in full (*full payment*), and 14 percentage points more likely to make payments on other arrears not included in the experiment (*other payments*).

The effect of the more impersonal methods (physical letter and email) is in line with previous tax compliance experiments that show that deterrence messages, if appropriately designed (personalized, and addressed and signed by a government official) work. The larger effect of the personal visits (particularly once we control for actual treatment in the next section) is in line with evidence regarding personal methods such as in the GOTV literature.

[24] The email and the personal visit are statistically different at the 10 percent level only for payment share.

IV.3 Taking into Account Non-Compliance with Assignment: TOT/LATE Estimations

As shown in Table 3, there were several sources of one-side non-compliance with the random assignment.[25] On the one hand, the Tax Agency double treated a small share of taxpayers (2 percent), with no cases of noncompliance in the control group. On the other hand, the agency did not have personnel-time to send all the letters and accomplish all the personal visits that had been scheduled, and some of the taxpayers the Agency tried to reach could not be located because either their physical or electronic address was wrong.[26] Consequently, about 38 percent of those assigned to the letter received a letter, 88 percent of those assigned to the email received an email, and 25 percent of those assigned to the personal visit were actually visited by a tax inspector.

The empirical exercises take these issues into account by looking not only at average ITT effects but also at TOT/LATE effects and by controlling for the fact that some people had already canceled their obligations (which a priori would bias the estimates downward), and other covariates that were not balanced during randomization.

[25] We had no contamination of the control group as the people in this group were removed from the sample the local agencies had access to for informing taxpayers. They still had access to the full set of people under treatment, which allowed them to pick and choose whether and how to treat them.

[26] While this number seems large, it is not uncommon even for countries with higher levels of compliance. For example, in fiscal year 2012, the IRS closed about 500 thousand cases (involving almost $7 billion of tax debt) because it could not locate delinquent taxpayers (Treasury General Inspector for Tax Administration, 2014).

Table 3. Compliance with the Experiment Design

	Treatment			
	Letter	**Email**	**Visit**	**Control group**
		Intent to treat		
Randomization assignment	5000	5000	4042	6776
		Non-compliance with assignment		
Attempted letter	4,394	0	0	0
Attempted email	41	4,982	30	0
Attempted inspection	116	11	1,270	0
		Failed treatments		
Failed letter	2,511	0	0	0
Failed email	1	584	3	0
Failed inspection	21	4	263	0
		Actual treatments		
Treated letter	1,883	0	0	0
Treated email	40	4,398	27	0
Treated inspection	95	7	1,007	0

Note: Each column presents the number of taxpayers that had been assigned to each treatment, the number the Agency attempted to contact, the number of times they failed, and finally the number actually treated. For example, out of 5000 assigned to a letter, the Agency only sent 4394 letters. Of those, only 1883 reached the taxpayers while 2511 were returned by the mail carriers because of problems locating the taxpayers. Source: Authors' calculations

In order to correct for this and to estimate the effect of the "revenue collection day" on the subset of effectively treated individuals, we instrument the actual treatment variable with the assignment to the treatment.[27] This way we can obtain complier average causal effect estimates. First-stage results are included in Online Appendix, while a summary of second-stage results is included in Table 4. Again, the top panel shows the results for the overall treatment and the bottom panel shows the results considering each treatment individually. Full regression tables are included in Online Appendix.

As expected, once we control by the fact that many taxpayers assigned to treatment were not contacted by the tax agency, the estimates are now substantially larger than before. For example, concentrating on the overall effect

[27] Unfortunately, we do not have inspector level information to control for inspector fixed effects.

of treatment (upper panel) shows that those treated had a 17-percentage points higher chance of making a payment (column [4]), and a 22-percentage points higher chance of making payments to cancel other liabilities (column [10]). In terms of money, they paid almost two-and-a-half times more than those in the control group (column [8]), which led to canceling about 15 percentage points more of the debt share (column [6]).

Moreover, the differences across mechanisms have become even more noticeable. The probability that people would make any payment (column [2]) has increased: 0.085 for letter, 0.17 for email, and 0.88 for personal visits; the probability that they would cancel the full amount of debt (column [4]) has also increased to 0.06, 0.13, and 0.65 respectively. The share of payments with respect to liabilities (column [6]) is also larger for the treatments than before: 0.07 for letter, 0.15 for email, and 0.73 for personal visits.[28] The same patterns of higher compliance exist also in terms of total payments and other payments, confirming once more the large spillover effect of the intervention.

[28] Some of the results could be underestimating the actual impact. First, while we know whether the letter was delivered, we have no information about whether the taxpayer actually read it or not. Second, while we know if an email was rejected by the server we have no information about whether the taxpayer actually received the email. Finally, while we assume that the inspector complied with the protocol, we do not have second-hand verification (and unfortunately we do not have inspector-level information to control for it).

Table 4. LATE (IV) Results

	Dependent variable									
	Paid		*Full payment*		*Payment share*		*Total payment (logs)*		*Other payments*	
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Overall Treatment	0.182***	0.174***	0.127***	0.130***	0.154***	0.152***	2.448***	2.346***	0.226***	0.215***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.11)	(0.11)	(0.01)	(0.01)
N	20,818	20,818	20,818	20,818	20,818	20,818	20,818	20,818	20,818	20,818
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
IV tests:										
LM test statistic for underidentification (Anderson or Kleibergen-Paap)	6819	6803	6819	6803	6819	6803	6819	6803	6819	6803
p-value of underidentification LM statistic	0	0	0	0	0	0	0	0	0	0
F statistic for weak identification (Cragg-Donald or Kleibergen-Paap)	13778	13715	13778	13715	13778	13715	13778	13715	13778	13715
Letter	0.092***	0.085***	0.055***	0.060***	0.068**	0.067**	1.307***	1.214***	0.302***	0.290***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.20)	(0.19)	(0.01)	(0.01)
Email	0.175***	0.169***	0.125***	0.127***	0.154***	0.152***	2.334***	2.250***	0.159***	0.152***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.10)	(0.10)	(0.01)	(0.01)
Personal Visit	0.897***	0.879***	0.630***	0.653***	0.727***	0.729***	12.071***	11.801***	0.899***	0.841***
	(0.07)	(0.07)	(0.05)	(0.05)	(0.08)	(0.08)	(0.92)	(0.90)	(0.07)	(0.06)
p-value Wald test eq. of coeff.	0	0	0	0	0	0	0	0	0	0
N	20,818	20,818	20,818	20,818	20,818	20,818	20,818	20,818	20,818	20,818
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
IV tests:										
LM test statistic for underidentification (Anderson or Kleibergen-Paap)	452.2	450.4	452.2	450.4	452.2	450.4	452.2	450.4	452.2	450.4
p-value of underidentification LM statistic	0	0	0	0	0	0	0	0	0	0
F statistic for weak identification (Cragg-Donald or Kleibergen-Paap)	165.2	164.3	165.2	164.3	165.2	164.3	165.2	164.3	165.2	164.3

Notes: Each row shows the regression coefficients and the standard error in parenthesis corresponding to the second stage of IV regression that include strata and district. Standard errors are robust. *** p<0.01, ** p<0.05, * p<0.1

The top section of the table shows the results for a regression that includes the overall treatment variable. The bottom section shows the results for regressions that include each treatment individually.

Even columns include Liabilities (in log), Taxpayer type (firm), Type of tax dummmies, Pre-payments (in logs), Wrong Information, and Overpayments as additional controls.

Source: Authors' calculations

Results in Ortega and Scartascini (2015), which look only at the effect of phone calls in a similar experimental setting, complement these results. Phone calls have an intermediate effect between the impersonal methods and the visit,

which is consistent with the framework in this paper. Personal methods have a larger impact (moral effect), but among those, the methods that are more costly to implement have a higher deterrence effect (the taxpayer internalizes a higher enforcement probability as shown in the model). Results are also in line with the evidence coming from the GOTV literature (summarized in Section 2), where personal canvassing has usually been more important than other mechanisms. For example, according to Imai (2005), personal canvassing was six times more effective than regular mail for getting people out to vote. Our results indicate that personal visit can be up to 10 times more effective than regular mail. The difference between these results could be explained at least in part by the deterrence component—which is not present in the GOTV case.[29]

What do these results tell us about the cost-benefit of each type of method? First, variable costs are different but relative low. The tax agency has calculated them to be about US$0 per email, US$0.50 per letter, and US$8 per personal visit. The average amount collected per attempted letter was around US$550, US$590 for the email, and more than US$2,000 for the attempted visits. Consequently, the net benefit of each intervention, considering only variable costs, favors the personal visits over the email and the letter (which is the least effective). This has implications for the optimal enforcement strategy. On the one hand, increasing the number of personal visits instead of relying on the impersonal methods seems to be highly cost-effective in spite of the fact that sending an inspector is much more costly than sending a letter. On the other hand, the relative effectiveness of the visit seems to be explained in part by how much taxpayers update the enforcement probability. Hence, a universal personal visits campaign may become less effective in the long run because taxpayers may not

[29] So far, the GOTV and related literatures have focused on moral/behavioral response to personal interactions. The results here show that rational reactions matter too and should be incorporated into the analysis (e.g., personal canvassing has an effect through personal interaction but it may also provide a signal that may affect the stakes for the individual in the electoral results).

internalize enforcement the same way (now, the probability of actual enforcement would be lower). However, restricting the use of impersonal methods (while being explicit about it) may increase their effect.

Overall, the set of results offers very important lessons. First, results are in line with the existing literature: enforcement matters. Contacting taxpayers in a personalized and detailed manner to inform them of their debts and the consequences of maintaining unpaid liabilities is effective for eliciting payments, at least in the short run. Second, the effect is not only significant for increasing payment of informed obligations but also because it generates substantial spillover effects. Third, because of high levels of non-compliance with assignment to the treatments, there are substantial differences between ITT and TOT/LATE estimates. This is a relevant finding that helps to put into perspective other results in the literature that have relied on ITT because they lacked information about who received and who did not receive treatment.[30] It also shows that there are plenty of gains to be made by simple strategies such as keeping databases up-to-date. Fourth, treatment effects based on individual-level behavior (the GOTV literature usually relies on district-level estimates) show that different delivery methods have substantially different effects on compliance. These effects may indicate potential unexploited gains in other policy areas too. Fifth, results are in line with the analytical framework. The more personalized the method, the higher the impact. Moreover, comparing the results in this paper with those in the GOTV and Ortega and Scartascini (2015) show that greater compliance is explained by both the effect of personal interaction and how much each individual updates the enforcement probability. Sixth, cost-benefit seems to favor increasing the number of personal visits, but the overall effect of a universalization of this

[30] This is the case, for example, in Castro and Scartascini (2015). Even though the authors took the precaution of sending the message on the property tax bill, and the address on the bills are associated with the addresses in the official property registry (which people have an incentive to keep up to date), there is a chance that some of the bills may have never reached the intended recipients.

strategy remains to be evaluated. Restricting the reach of impersonal methods may have a payoff too. Finally, electronic methods seem to be more effective than traditional letters. The exact mechanism behind this result may warrant further research.[31]

IV.4 Heterogeneous Results

Are the results different for different types of taxpayers? In order to check for potential heterogeneous effects we have interacted the treatments with the control variables that proxy observable differences across taxpayers. Table 5 presents a short summary of the results for a subset of the dependent variables for the overall treatment. Complete results for each treatment are included in Online Appendix.

First, taxpayers with standing liabilities on the income tax and VAT seem to react more to the treatments than those who owe wealth taxes. This pattern repeats for the individual treatments. Second, those with medium level of debts seem to react slightly more than those with low and higher debt. This overall effect does not hold up for every treatment. For example, people with medium debt are less likely to comply when they receive an email. Third, in general there seems to be little difference between legal individuals and natural persons. At the individual treatment level, legal individuals seem to react more than natural people when visited by an inspector. Finally, how old the debt is seems to be unrelated to treatment response.

From these results, it is difficult to elicit the exact mechanisms at work. For example, the results regarding the differences between firms and individuals may be due to a higher response of legal entities to threats, but it may also be explained by higher cash constraints for individuals. Similarly, those with high

[31] One possibility may be salience. The agency had been moving many of its transactions online, so the email may have had a relatively higher salience, which may not export easily to other places. Additionally, given the fact that payments can be made online, the act of paying may have been more spontaneous than after receiving a letter (the person was already sitting at the computer).

debts may be reacting less, either because they are larger entities and have a better ability to dodge the law or because they have accumulated so much debt that it is more difficult for them to react in the short term. Unfortunately, we do not have information that could help us to differentiate between these mechanisms. Regarding the type of tax, we believe that financial constraints play a major role here for the much lower response from those who owe wealth taxes than those who owed other taxes. Wealth taxes affect an asset, which may be illiquid, while the VAT and income taxes tax the flow of revenues.

Table 5. Heterogeneous Effects

Dependent variables:	Dependent variable				
	Paid	**Total payments (in logs)**	**Payment share**	**Full payment**	**Other payments**
			Type of Tax		
Overall treatment	-0.048*	-0.715*	-0.035	-0.043**	2.352***
	(0.03)	(0.38)	(0.05)	(0.02)	(0.34)
Treatment*Income Tax	0.227***	2.973***	0.259***	0.254***	-0.238
	(0.03)	(0.48)	(0.07)	(0.03)	(0.43)
Treatment*VAT	0.258***	3.598***	0.197***	0.179***	0.975***
	(0.03)	(0.41)	(0.06)	(0.02)	(0.36)
			Debt Size		
Overall treatment	0.170***	2.295***	0.170***	0.129***	2.452***
	(0.02)	(0.22)	(0.03)	(0.01)	(0.20)
Treatment*Medium Debt	0.017	-0.258	0.003	0.073***	-0.033
	(0.02)	(0.33)	(0.05)	(0.02)	(0.30)
Treatment*High Debt	-0.001	0.373	-0.052	-0.058***	1.470***
	(0.02)	(0.31)	(0.04)	(0.02)	(0.28)
			Firms		
Overall treatment	0.166***	2.198***	0.149***	0.125***	2.314***
	(0.02)	(0.21)	(0.03)	(0.01)	(0.13)
Overal treatment*Firms	0.014	0.244	0.004	0.007	0.543
	(0.02)	(0.27)	(0.04)	(0.02)	(0.52)
			Debt Age		
Overall treatment	0.172***	2.314***	0.147***	0.125***	0.216***
	(0.01)	(0.13)	(0.02)	(0.01)	(0.01)
Overall treatment *liability age is low	0.036	0.543	0.059	0.061**	-0.008
	(0.04)	(0.52)	(0.07)	(0.03)	(0.03)

Notes: Standard errors in parenthesis. *** denotes significance at the 1% level; ** at 5%; * at 1%
These estimations correspond to the second stage of IV regressions with the following controls: block dummies, Type of Tax, Taxpayer type (firms), Actual liabilities (in logs), Pre-payments (in logs), Wrong information, Negative debt, and distric-specific dummies
The endogenous variable, actual treatments, have been instrumented with the assignment to treatment.
Source: Authors' calculations

IV.5 Explaining Assignment Non-Compliance

As we have mentioned, the tax authority could not find many of the taxpayers assigned to treatment. In many cases, these taxpayers may have moved or their email account may have been deactivated. One important question is whether there are common patterns across these groups. For example, if non-compliance is explained by the size or number of outstanding obligations—i.e., those with more debt have an incentive to provide false information—then the interpretation of the results applies to those who received treatment and not to the entire experimental population.

We present a full analysis in Online Appendix evaluating the characteristics of those who could not be located, and the characteristics of those the agency decided to visit. From the empirical analysis, while we recognize that our results are local, we do not see any specific biases in terms of the sample we are using. First, it does not seem to be the case that the wrong addresses have been the result of a conscious decision by the taxpayers to avoid prosecution. Second, even if there is a slightly higher chance of being visited according to the size of the debt, results are not economically meaningful (someone in the 90^{th} percentile would have had approximately a 2 percent higher probability of being audited than somebody in the 50^{th} percentile.)

V. Conclusions

The literature has shown that sending messages has an effect on compliance, and that different messages in terms of both the content (e.g., deterrence, moral) and the characteristics of the messages (e.g., whether they are signed by a tax agency authority or not) have different impacts. Evaluating the delivery mechanism for

the messages, which has been common in related literatures, has been absent in the tax compliance literature.

The results in this paper show that campaigns that inform taxpayers regarding pending liabilities are a good mechanism for increasing compliance. In the case of the campaign run by the Tax Agency in Colombia, the evidence indicates that running the campaign increased compliance significantly. The agency collected two-and-a-half times the amount it would have collected if it had done nothing, which helped them recover about one fourth of the outstanding debt of those contacted (around US$8 million PPP). Regression results show that the difference between doing nothing and running the campaign increases the probability of receiving a payment by 10 percentage points (ITT results) and by almost 20 points when we consider only those who were effectively treated (TOT/LATE results). Moreover, there are large spillover effects. The campaign increased not only cancellation of pending liabilities reported by the agency to taxpayers but also the payment of other pending obligations.

Of course, not every method for contacting the taxpayer works the same. On the one hand, each method has a different impact on taxpayers' perceptions of the severity and consequences of non-compliance: receiving a visit from an inspector seems to be more effective than the impersonal methods. On the other hand, databases are not always up-to-date, and some methods require more human effort than others to reach the taxpayer; hence, emails seem to be more effective than letters in eliciting payments. Thus, according to the LATE estimates, the probability of making any payment is 8 percentage points higher for the letter, 17 points for the email, and about 87 percentage points for the personal visit. That is, almost every person who received a visit decided to make some sort of payment to the Agency, which implied recovering around 70 percent of the amount owed by them.

The policy implications of these results are clear. There are plenty of gains to be made by Tax Agencies by contacting the taxpayers regarding their standing liabilities, and even more so if they keep a clean and up-to-date contact information database. Results indicate that having a valid physical and electronic address for each taxpayer could have a large payoff; in the restricted sample we used, having a valid address might had implied doubling collection to about US$8 million PPP. Consequently, implementing a strategy that ensures that each taxpayer has a valid and working account may be as important as many more sophisticated and costly enforcement strategies that have been tried in the past. A simple cost-benefit analysis indicates that the net benefit of each letter and email sent was about the same—but much lower than the benefit of each personal visit attempted. This finding has implications for the optimal enforcement strategy. On the one hand, increasing the number of personal visits instead of relying on the impersonal methods seems to be cost-effective. On the other hand, the relative effectiveness of the visit seems to be explained in part by how much taxpayers update the enforcement probability. Hence, a universal personal visits campaign may become less effective if taxpayers realize it is being implemented. On the contrary, an email campaign that indicates that the taxpayer is being targeted with a small and selected group of taxpayers could become very effective.

In terms of academic implications, the results raise the bar for future field experiments and open up new venues of research. First, future work could incorporate the idea that not only the role of messages should be evaluated but also the delivery mechanisms. Hence, to isolate each effect, it may make sense to consider randomizing both the message and the method. In particular, some types of messages may be more effective when delivered by some methods than by others. For example, moral suasion messages may be relatively more effective when delivered by an individual in a personalized manner than in an impersonal

method as a letter, which has usually been the norm. This is an empirical question that should be extended to other policy areas as well.

Second, future exercises should incorporate the variation of methods to test the effect of the different delivery mechanisms on standard compliance settings instead of looking only at tax delinquencies as we do here. Third, even though comparing the results in this paper to the GOTV results provides a first approximation to evaluating how much of the effect of the personal visits is due to the moral effect and how much to the update in the enforcement probability, it would still be necessary to test which of those effects is dominant. Future experiments could include explicit messages showing that the number of taxpayers being contacted is fixed and also the contact method to reach them. This way, any difference between the personal visit and the impersonal method would be due to moral considerations.

Fourth, the paper has shown that spillover effects can be substantial. In the case of this experiment, the messages elicited additional payments. It could be the case in other contexts, particularly in the context of enforcement of taxes with self-declaration, taxpayers may substitute across taxes and compensate by paying what the government asks but reducing the amount declared in other taxes (or to other authorities). Whenever possible, studies should incorporate evaluating spillover effects explicitly into the research strategy.

Finally, the findings have implications beyond the deterrence exercise we have developed here. They highlight that the mechanisms through which policies are informed and publicized should not be neglected from the economic policy debate.

References

Aguirre, B., and F.F. Rocha. 2010. "Trust and Tax Morale in Latin American and Caribbean Countries." Paper presented at the 14th Annual Conference of the International Society for New Institutional Economics, June 17–19, Stirling, Scotland, United Kingdom.

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

Alm, J., B. Jackson and M. McKee. 1992. "Institutional Uncertainty and Taxpayer Compliance." *American Economic Review* 82(4): 1018–26.

Alm, J., and J. Martínez-Vázquez. 2007. "Tax Morale and Tax Evasion in Latin America." International Studies Program Working Paper 07-04. Atlanta, United States; Georgia State University.

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

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

Andreoni, J., B. Erard and J. Feinstein. 1998. "Tax Compliance." *Journal of Economic Literature* 36(2): 818–60.

Arceneaux, K. 2007. "I'm Asking for Your Support: The Effect of Personally Delivered Campaign Messages on Voting Decisions and Opinion Formation." *Quarterly Journal of Political Science* 2(1): 43-65.

Arceneaux, K., and D. Nickerson. 2006. "Even if You Have Nothing Nice to Say, Go Ahead and Say It: Two Field Experiments Testing Negative Campaign Tactics." Unpublished manuscript. Prepared for presentation at the 2005 Meeting of the American Political Science Association, September 1-4, Washington, DC, United States.

Ariel, B. 2012. “Deterrence and Moral Persuasion Effects on Corporate Tax Compliance: Findings from a Randomized Controlled Trial.” *Criminology* 50(1): 27–69.

Barone, G., and S. Mocetti. 2009. “Tax Morale and Public Spending Inefficiency.” Working Paper 732. Rome, Italy: Banca d’Italia.

Barton, J., M. Castillo and R. Petrie. 2014. “What Persuades Voters? A Field Experiment on Political Campaigning.” *Economic Journal* 124(574): F293–F326.

Bayer, R.C. 2006. “Moral Constraints and Evasion of Income Tax.” *IUP Journal of Public Finance* 4(1): 7–31.

Behavioral Insights Team (BIT). 2012 “Applying Behavioural Insights to Reduce Fraud, Error and Debt.” London, United Kingdom: Cabinet Office. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60539/BIT_FraudErrorDebt_accessible.pdf

Bergman, M., and A. Nevarez. 2006. “Do Audits Enhance Compliance? An Empirical Assessment of VAT Enforcement.” *National Tax Journal* 59(4): 817–32.

Bernheim, B.D., and A. Rangel. 2007. “Behavioral Public Economics, Welfare and Policy Analysis with Nonstandard Decision-Makers.” In: P. Diamond and H. Vartiainen, editors. *Behavioral Economics and Its Applications*. Princeton, United States: Princeton University Press.

Bernheim, B.D., and A. Rangel. 2009. “Beyond Revealed Preference, Choice-Theoretic Foundations for Behavioral Welfare Economics.” *Quarterly Journal of Economics* 124(1): 51–104.

Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir and J. Zinman. 2010. “What’s Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment.” *Quarterly Journal of Economics* 125(1): 263–305.

Blumenthal, M., C. Christian and J. Slemrod. 2001. "Do Normative Appeals Affect Tax Compliance? Evidence from a Controlled Experiment in Minnesota." *National Tax Journal* 54(1): 125–38.

Bruhn, M., and D. McKenzie. 2009. "In Pursuit of Balance, Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics* 1(4): 200–32.

Card, D., S. DellaVigna and U. Mamendier. 2011. "The Role of Theory in Field Experiments." NBER Working Paper 17047. Cambridge, United States: National Bureau of Economic Research.

Castro, L., and C. Scartascini. 2015. "Tax Compliance and Enforcement in the Pampas: Evidence from a Field Experiment." *Journal of Economic Behavior and Organization* 116: 65–82.

Chetty, R., A. Looney and K. Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99(4): 1145-1177.

Chirico, M., R, Inman, C, Loeffler, J, MacDonald, and H, Sieg. 2015. "An Experimental Evaluation of Notification Strategies to Increase Property Tax Compliance: Free-Riding in the City of Brotherly Love." In: J.R. Brown, editor. *Tax Policy and the Economy*. Volume 30. Chicago, United States: National Bureau of Economic Research and University of Chicago Press.

Coleman, S. 1996. "The Minnesota Income Tax Compliance Experiment: State Tax Results." MPRA Paper 4827. Munich, Germany: Munich University Library.

Coleman, S., 2007. "The Minnesota Income Tax Compliance Experiment, Replication of the Social Norms Experiment." MPRA Paper 5820. Munich, Germany: Munich University Library.

Congdon, W.J., J.R. Kling and S. Mullainathan. 2011. "Policy and Choice, Public Finance through the Lens of Behavioral Economics." Washington, DC, United States: Brookings Institution Press.

Corbacho, A., V. Fretes Cibils and E. Lora, editors. 2013. "More than Revenue: Taxation as a Development Tool." Development in the Americas Report. Washington, DC, and New York, United States: Inter-American Development Bank and Palgrave Macmillan.

Cowell, F.A., and J.P.F. Gordon. 1988. "Unwillingness to Pay, Tax Evasion and Public Good Provision." *Journal of Public Economics* 36(3): 305–21.

Cummings, R.G. et al. 2005. "Effects of Tax Morale on Tax Compliance, Experimental and Survey Evidence." CREMA Working Paper 2005-29. Basel, Switzerland: Center for Research in Economics, Management and the Arts.

Dale, A., and A. Strauss. 2009. "Don't Forget to Vote: Text-Messages Reminders as a Mobilization Tool." *American Journal of Political Science* 53(4): 787-804

Dell'Anno, R. 2009. "Tax Evasion, Tax Morale and Policy Maker's Effectiveness." *Journal of Socio-Economics* 38(6): 988–97.

DellaVigna, S., J.A. List and U. Malmendier 2009. "Testing for Altruism and Social Pressure in Charitable Giving." NBER Working Paper 15629. Cambridge, United States: National Bureau of Economic Research.

Del Carpio, L. 2014. "Are the Neighbors Cheating? "Evidence from a Social Norm Experiment on Property Taxes in Peru." Princeton University Manuscript.
http://scholar.princeton.edu/sites/default/files/Are_the_neighbors_cheating_Apr2014_0.pdf

Dhami, S., and A. al-Nowaihi. 2007. "Why Do People Pay Taxes? Prospect Theory versus Expected Utility Theory." *Journal of Economic Behavior and Organization* 64(1): 171–92.

Dhami, S., and A. al-Nowaihi. 2010. "Optimal Taxation in the Presence of Tax Evasion: Expected Utility versus Prospect Theory." *Journal of Economic Behavior and Organization* 75(2): 313–37.

Dwenger N., H. Kleven, I. Rasul, and J. Rincke. 2014. "Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany." Unpublished manuscript. Available at: http://www.tax.mpg.de/files/pdf3/dwenger-kleven-rasul-rincke_oct2014_full1.pdf

Eisenhauer, J.G. 2008. "Ethical Preferences, Risk Aversion, Taxpayer Behavior." *Journal of Socio-Economics* 37(1): 45–63.

Fellner, G., R. Sausgruber and C. Traxler. 2013. "Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information." *Journal of the European Economic Association* 11(3): 634–60.

Fortin, B., G. Lacroix and M.C. Villeval. 2007. "Tax Evasion and Social Interactions." *Journal of Public Economics* 91(11–12): 2089–2112.

Freeman, R.B. 1997. "Working for Nothing: The Supply of Volunteer Labor." Journal of Labor Economics 15(1):S140-66.

Frey, B.S., and B. Torgler. 2007. "Tax Morale and Conditional Cooperation." *Journal of Comparative Economics* 35(1): 136–59.

Fryer, R.G., Jr., 2013. "Information and Student Achievement, Evidence from a Cellular Phone Experiment." NBER Working Paper 19113. Cambridge, United States: National Bureau of Economic Research.

Fujiwara, T., and L. Wantchekon. 2013. "Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin." *American Economic Journal: Applied Economics* 5(4): 241–255

Gangl, K. et al. 2014. “Effects of Supervision on Tax Compliance: Evidence from a Field Experiment in Austria.” *Economic Letters* 123: 378-382.

Gerber, A.S., and D.P. Green. 2000. “The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment.” *American Political Science Review* 94: 653-63.

Gordon, J.P.F. 1989. “Individual Morality and Reputation Costs as Deterrents to Tax Evasion.” *European Economic Review* 33(4): 797–805.

Green, A., and D. Gerber. 2008. “Get Out the Vote: How to Increase Voter Turnout.” Washington, DC, United States: Brookings Institution Press.

Green, D., and D. Karlan. 2006. “Effects of Robotic Calls on Voter Mobilization.” Unpublished manuscript. New Haven, United States: Yale University,Institution for Social and Policy Studies.

Hallsworth, M., J. List and I.V. Metcalfe. 2014. “The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance.” NBER Working Paper 20007. Cambridge, United States: National Bureau of Economic Research.

Hallsworth, M. 2014. “The Use of Field Experiments to Increase Tax Compliance.” *Oxford Review of Economic Policy* 30(4): 658-679.

Harbaugh, W.T. 1998. “What Do Donations Buy? A Model of Philanthropy Based on Prestige and Warm Glow.” *Journal of Public Economics* 67: 269-284.

Hashimzade, N., G.D. Myles and B. Tran-Nam. 2010. “New Approaches to the Economics of Tax Evasion.” Exeter, United Kingdom: University of Exeter. http://people.exeter.ac.uk/gdmyles/papers/pdfs/Non-euEva.pdf

Hashimzade, N., G.D. Myles and B. Tran-Nam. 2012. “Applications of Behavioural Economics to Tax Evasion.” *Journal of Economic Surveys*. Published online, 24 May 2012. DOI: 10.1111/j.1467-6419.2012.00733.x

Hasseldine, J., P.A. Hite, S. James and M. Toumi. 2007. "Persuasive Communication: Tax Compliance Enforcement Strategies for Sole Proprietors." *Contemporary Accounting Research* 24(1): 171–194

Haynes, L., D.P. Green, R. Gallagher, P. Jhon and D.J. Torgerson. 2013. "Collection of Delinquent Fines: An Adaptive Randomized Trial to Assess the Effectiveness of Alternative Text Messages." *Journal of Policy Analysis and Management* 32(4): 718–730.

Imai, K. 2005. "Do Get-Out-the-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments." *American Political Science Review* 99: 283–300.

Inter-American Development Bank (IDB). 2011. "Development Connections. Unveiling the Impact of New Information Technologies." New York, United States: Palgrave Macmillan.

Khan, A.Q., A.I. Khwaja and B.A. Olken. 2014. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors." NBER Working Paper 20627. Cambridge, United States: National Bureau of Economic Research.

Kendall, C., T. Nannicini and F. Trebbi. 2013. "How Do Voters Respond to Information? Evidence from a Randomized Campaign." NBER Working Paper 18986. Cambridge, United States: National Bureau of Economic Research.

Kessler, J.B., and C.Y. Zhang. 2014. "Behavioral Economics and Health." Paper for *Oxford Textbook of Public Health*. Available at: http://assets.wharton.upenn.edu/~czhan/KesslerZhang_BehavioralEconomicsHealth.pdf

Kim, Y., 2003. "Income Distribution and Equilibrium Multiplicity in a Stigma-Based Model of Tax Evasion." *Journal of Public Economics* 87: 1591–1616.

Kleven, H.J., M. Knudsen, C. Kreiner, S. Pedersen, and E. Saez. 2011. "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark." *Econometrica* 79(3): 651–92.

Lacetera, N., and M. Macis. 2010. "Social Image Concerns and Prosocial Behavior: Field Evidence from a Nonlinear Incentive Scheme." *Journal of Economic Behavior and Organization* 76(2): 225-37.

Landry, C., A. Lange, J. A. List, M. K. Price and N. G. Rupp. 2006. "Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment." *Quarterly Journal of Economics* 121(2): 747-782.

Luttmer, E.F.P., and M. Singhal. 2014. "Tax Morale." *Journal of Economic Perspectives* 28(4): 149-168.

Mascagni, G. 2014. "A Review of Tax Experiments: From the Lab to the Field." Evidence Report 97. Essex, United Kingdom: Institute of Development Studies.

Meer, J., and H.S. Rosen. 2011. "The ABCs of Charitable Solicitation." *Journal of Public Economics* 95(5-6):363-71.

Murphy, K. 2004. "The Role of Trust in Nurturing Compliance: A Study of Accused Tax Avoiders." *Law and Human Behavior* 28(2): 187–209.

Myles, G.D., and R.A. Naylor. 1996. "A Model of Tax Evasion with Group Conformity and Social Customs." *European Journal of Political Economy* 12(1): 49–66.

Nickerson, D. 2006a. "Volunteer Phone Calls Can Increase Turnout." *American Politics Research* 34(3): 271-292.

Nickerson, D. 2006b. "Demobilized by e-Mobilization: Evidence from Thirteen Field Experiments." Notre Dame, United States: University of Notre Dame, Department of Political Science. Unpublished Manuscript.

Ortega, D., and P. Sanguinetti. 2013. "Deterrence and Reciprocity Effects on Tax Compliance: Experimental Evidence from Venezuela." CAF Working Paper 2013/08. Caracas, Venezuela: CAF.

Ortega, D., and C. Scartascini. 2015. "Who's Calling? The Effect of Phone Calls as a Deterrence Mechanism." Washington, DC, United States: Inter-American Development Bank. Mimeographed document.

Pérez-Truglia, R., and U. Troiano. 2015. "Shaming Tax Delinquents: Theory and Evidence from a Field Experiment in the United States." NBER Working Paper 21264. Cambridge, United States: National Bureau of Economic Research.

Pomeranz, D. 2013. "No Taxation without Information, Deterrence and Self-Enforcement in the Value Added Tax." NBER Working Paper 19199. Cambridge, United States: National Bureau of Economic Research.

Ramírez, R. 2005. "Giving Voice to Latino Voters: A Field Experiment on the Effectiveness of a National Nonpartisan Mobilization Effort." *Annals of the American Academy of Political and Social Science* 601, 66-84.

Rawlings, G., 2004. "Cultural Narratives of Taxation and Citizenship, Fairness, Groups and Globalisation." Working Paper 52. Canberra, Australia: Australian National University: Centre for Tax System Integrity.

Rogers, T., A.S. Gerber and C.R. Fox. 2012. "Rethinking Why People Vote: Voting as Dynamic Social Expression." In: E. Shafir, editor. *Behavioral Foundations of Policy*. Princeton, United States: Princeton University Press.

Sandmo, A. 2005. "The Theory of Tax Evasion: A Retrospective View." *National Tax Journal* 58(4): 643–64.

Scholz, J.T., and M. Lubell. 1998. "Trust and Taxpaying, Testing the Heuristic Approach to Collective Action." *American Journal of Political Science* 42(2): 398–417.

Scholz, J.T., and N. Pinney. 1995. “Duty, Fear, Tax Compliance: The Heuristic Basis of Citizenship Behavior.” *American Journal of Political Science* 39(2): 490–512.

Shaw, D.R., D.P. Green, J.G. Gimpel and A. S. Gerbe. 2012. Do Robotic Calls from Credible Sources Influence Voter Turnout or Vote Choice? Evidence from a Randomized Field Experiment.” *Journal of Political Marketing* 11(4): 231-245.

Slemrod, J., M. Blumenthal and C. Christian. 2001. “Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota.” *Journal of Public Economics* 79(3): 455–83.

Snow, A., and R. Warren. 2005. “Ambiguity about Audit Probability, Tax Compliance, and Taxpayer Welfare.” *Economic Inquiry* 43(4): 865-871.

Stollwerk, A.F. 2006. “Does E-mail Affect Voter Turnout? An Experimental Study of the New York City 2005 Election.” New Haven, United States: Yale University, Institution for Social and Policy Studies. Unpublished manuscript.

Thaler, R.H., and C.R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, Happiness*. New Haven, United States: Yale University Press.

Torgler, B. 2003. “Tax Morale, Theory and Empirical Analysis of Tax Compliance.” Basel, Switzerland: University of Basel. Ph.D. dissertation.

Torgler, B. 2004. “Moral Suasion, An Alternative Tax Policy Strategy? Evidence from a Controlled Field Experiment in Switzerland.” Economics of Governance 5(3): 235–53.

Torgler, B. 2005. “Tax Morale in Latin America.” *Public Choice* 122(1/2): 133–57.

Torgler, B. 2013. "A Field Experiment in Moral Suasion and Tax Compliance Focusing on Underdeclaration and Overdeduction." *Public Finance Analysis* 69(4): 393-411.

Torgler, B., A. Macintyre and M. Schaffner. 2008. "Causes and Consequences of Tax Morale: An Empirical Investigation." *Economic Analysis and Policy* 38(2): 313–39.

Traxler, C. 2010. "Social Norms and Conditional Cooperative Taxpayers." *European Journal of Political Economy* 26(1): 89–103.

Traxler, C., and J. Winter. 2012. "Survey Evidence on Conditional Norm Enforcement." *European Journal of Political Economy* 28(3): 390–98.

Treasury General Inspector for Tax Administration, 2014. "Delinquent Taxes May Not Be Collected Because Required Research Was Not Always Completed Prior to Closing Some Cases As Currently Not Collectible." Reference Number: 2014-30-052. Department of the Treasury. Washington, DC, United States: U.S. http://www.treasury.gov/tigta/auditreports/2014reports/201430052fr.pdf

Wenzel, M. 2002. "Altering Norm Perceptions to Increase Tax Compliance." Working Paper 38. Canberra, Australia: Australian National University, Centre for Tax System Integrity.

Wenzel, M. 2007. "The Multiplicity of Taxpayer Identities and Their Implications for Tax Ethics." *Law and Policy* 29(1): 31–50.

Yaniv, G. 1999. "Tax Compliance and Advance Tax Payments: A Prospect Theory Analysis." *National Tax Journal* 52(4): 753–64.

Yitzhaki, S. 1974. "A Note on Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics* 3(2): 201–02.

Appendix. Description of Variables

Randomization was performed according to taxpayer's liabilities, which was the information to be provided in the messages, in six blocks according to size of debt and maturity. As can be observed in the table, samples balance on that variable. Unfortunately, they don't balance in some of the other covariates; we include them as controls in the empirical analysis.

Table A.1. Random Assignment to Treatment

		Difference w.r.t. control (coeff and s.e.)								
	Average	Overall	Individual Treatments			p-value Wald test equality coefficients				Sample
	and s.d.	treatment	Letter	Email	Visit	[3]=[4]	[3]=[5]	[4]=[5]	[3]=[4]=[5]	size
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Liabilities (in millions)	4.440	0.026	-0.024	0.019	0.172	0.723	0.135	0.277	0.32	20,818
	(7.731)	(0.098)	(0.120)	(0.113)	(0.144)					
Liabilities (in logs)	13.998	0.009	0.001	0.01	0.023	0.524	0.195	0.489	0.425	20,818
	(1.820)	(0.012)	(0.015)	(0.014)	(0.019)					
Number of debts	1.753	0.015	0.001	0.031	0.002	0.267	0.981	0.345	0.491	20,818
	(1.421)	(0.022)	(0.027)	(0.025)	(0.0316)					
Tax (Wealth)	0.105	-0.004	-0.001	-0.012**	0.007	0.067	0.273	0.011	0.03	20,818
	(0.307)	(0.005)	(0.006)	(0.006)	(0.00758)					
Tax (Income Tax)	0.229	-0.002	0.002	-0.007	0.001	0.27	0.902	0.437	0.518	20,818
	(0.420)	(0.006)	(0.008)	(0.008)	(0.00986)					
Tax (VAT)	0.666	0.007	-0.001	0.019**	-0.008	0.032	0.55	0.02	0.032	20,818
	(472)	(0.007)	(0.009)	(0.009)	(0.0113)					
Taxpayer type (firms)	0.616	0.055***	0.046***	0.049***	0.095***	0.764	0	0	0	20,818
	(0.486)	(0.008)	(0.009)	(0.009)	(0.0116)					

Notes: Each row shows statistics for a different variable. . Column [1] shows the sample average and the standard deviation in parenthesis
Columns [2] shows the regression coefficient and the standard error in parenthesis corresponding to an OLS regression that includes controls for strata and district. Standard errors are robust. *** p<0.01, ** p<0.05, * p<0.1
Columns [3]-[5] shows the regression coefficients and the standard errors in parenthesis corresponding to an OLS regression that includes controls for strata and district. Standard errors are robust. *** p<0.01, ** p<0.05, * p<0.1
Columns [6]-[9] shows the p-value of a test of equality of coefficients. Column [10] shows the sample size.
Source: Authors' calculations

Table A.2. First-Stage Regression Table

	Dependent variables			
	Overall Treatment		Treated Letter	
VARIABLES	(1)	(2)	(3)	(4)
Assignment to Treatment	0.600***	0.601***		
	(0.01)	(0.01)		
Assignment to Letter			0.400***	0.400***
			(0.01)	(0.01)
Assignment to Email			0.004***	0.004***
			(0.00)	(0.00)
Assignment to Inspection			0.053***	0.052***
			(0.00)	(0.00)
N	20,818	20,818	20,818	20,818
Controls	No	Yes	No	Yes
LM test statistic for underidentification (Anderson or Kleibergen-Paap)	6819	6803	6819	6803
p-value of underidentification LM statistic	0	0	0	0
F statistic for weak identification (Cragg-Donald or Kleibergen-Paap)	13778	13715	13778	13715
	Dependent variables			
	Treated Email		Treated Inspection	
	(5)	(6)	(7)	(8)
Assignment to Letter	0.008***	0.009***	0.004**	0.004**
	(0.00)	(0.00)	(0.00)	(0.00)
Assignment to Email	0.879***	0.879***	-0.001	-0.001
	(0.00)	(0.00)	(0.00)	(0.00)
Assignment to Inspection	0.010***	0.011***	0.145***	0.144***
	(0.00)	(0.00)	(0.01)	(0.01)
N	20,818	20,818	20,818	20,818
Controls	No	Yes	No	Yes
LM test statistic for underidentification (Anderson or Kleibergen-Paap)	452.2	450.4	452.2	450.4
p-value of underidentification LM statistic	0	0	0	0
F statistic for weak identification (Cragg-Donald or Kleibergen-Paap)	165.2	164.3	165.2	164.3
N	20,818	20,818	20,818	20,818
Controls	No	Yes	No	Yes

Notes: Each row shows the regression coefficients and the standard error in parenthesis correspondin to the First stage of IV regression that include strata and district. Standard errors are robust.
*** p<0.01, ** p<0.05, * p<0.1
Even columns include Liabilities (in log), Taxpayer type (firm), Type of tax dummies, Pre-payments (in logs), Wrong Information, and Overpayments as additional controls.
Source: Authors' calculations