NBER WORKING PAPER SERIES

TAX AUDITS AS SCARECROWS:
EVIDENCE FROM A LARGE-SCALE FIELD EXPERIMENT

Marcelo L. Bérgolo
Rodrigo Ceni
Guillermo Cruces
Matias Giaccobasso
Ricardo Perez-Truglia

Working Paper 23631
http://www.nber.org/papers/w23631

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2017, Revised May 2019

We thank the Uruguay’s national tax administration (Dirección General Impositiva) for their collaboration. We thank Gustavo Gonzalez for his support, without which this research would not have been possible. We thank Joel Slemrod for his valuable feedback, as well as that of seminar participants at University of Michigan, University of California San Diego, Dartmouth University, Universidad Di Tella, Universidad de la Republica, Universidad de Santiago de Chile, Universidad Católica de Chile, CAF, Banco Central del Uruguay, the 2017 NBER Public Economics Fall Meeting, the 2017 RIDGE Public Economics Conference, the 2017 Zurich Center for Economic Development Conference, the 2017 Advances with Field Experiments Conference, the 2018 PacDev Conference, the 2018 AEA Annual Meetings, the 2018 LAGV Conference, and the 2018 IIPF Annual Congress. This project benefited from funding by CEF, CEDLAS-UNLP and IDRC. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Marcelo L. Bérgolo, Rodrigo Ceni, Guillermo Cruces, Matias Giaccobasso, and Ricardo Perez-Truglia. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment
Marcelo L. Bérgolo, Rodrigo Ceni, Guillermo Cruces, Matias Giaccobasso, and Ricardo Perez-Truglia
NBER Working Paper No. 23631
July 2017, Revised May 2019
JEL No. C93,H26,K42

ABSTRACT

The canonical model of Allingham and Sandmo (1972) predicts that firms evade taxes by optimally trading off between the costs and benefits of evasion. However, there is no direct evidence that firms react to audits in this way. We conducted a large-scale field experiment in collaboration with Uruguay's tax authority to address this question. We sent letters to 20,440 small- and medium-sized firms that collectively paid more than 200 million dollars in taxes per year. Our letters provided exogenous yet nondeceptive signals about key inputs for their evasion decisions, such as audit probabilities and penalty rates. We measured the effect of these signals on their subsequent perceptions about the auditing process, based on survey data, as well as on the actual taxes paid, based on administrative data. We find that providing information about audits had a significant effect on tax compliance but in a manner that was inconsistent with Allingham and Sandmo (1972). Our findings are consistent with an alternative model, risk-as-feelings, in which messages about audits generate fear and induce probability neglect. According to this model, audits may deter tax evasion in the same way that scarecrows frighten off birds.

Marcelo L. Bérgolo
Instituto de Economia (IECON)
Universidad de La Republica
1926 Gonzalo Ramirez
Montevideo, Uruguay
mbergolo@iecon.ccee.edu.uy

Rodrigo Ceni
Instituto de Economia (IECON)
Universidad de La Republica
1375 Joaquin Requena
Montevideo, Uruguay
rceni@iecon.ccee.edu.uy

Guillermo Cruces
CEDLAS
Univesidad Nacional de La Plata
Calle 6 entre 47 y 48
La Plata, Argentina
gcruces@cedlas.org

Matias Giaccobasso
Anderson School of Management - UCLA
110 Westwood Plaza, C 3.10
Los Angeles, CA 90077
mgiaccobasso@ad.ucla.edu

Ricardo Perez-Truglia
Anderson School of Management
University of California, Los Angeles
110 Westwood Plaza
Los Angeles, CA 90095
and NBER
ricardo.truglia@anderson.ucla.edu

An online appendix is available at http://www.nber.org/data-appendix/w23631

1 Introduction

Tax audits have been standard tools of most tax administrations throughout history. Audits increase tax revenues directly, because firms caught evading must pay taxes on the hidden income and corresponding penalties. However, with the exception of large taxpayers, these direct revenues are insufficient to make audits cost-effective. Audits play a central role in the deterrence paradigm of tax evasion: the threat of being audited in the future, of being caught evading and having to pay penalties deters firms from evading taxes in the present.

Audits may be useful to foster tax compliance, but there is no direct evidence on how firms react to audits. The Allingham and Sandmo (1972) model (hereafter, A&S) is the canonical model of tax evasion in economics. This model is an application of Becker (1968), in which selfish individuals choose whether to engage in criminal activities based on the tradeoff between expected costs and benefits. In *A&S*, firms choose the optimal amount of income to hide from the tax authority so that the marginal benefits (i.e., the lower tax burden) equal the marginal costs (i.e., the penalties if caught). Whether firms in the real world react to audits in a calculated manner, as in *A&S*, is still a source of debate (Alm et al., 1992; Dhami and al Nowaihi, 2007; Luttmer and Singhal, 2014; Slemrod, 2018). In this study, we provide direct tests of the *A&S* model based on a high-stakes, large-scale field experiment.

We study a context in which firms should be most attentive to the threat of being audited: small- and medium-sized firms that are subject to the Value Added Tax (VAT). For other sources of taxable income, such as wage income, tax agencies can use third-party reporting to detect evasion automatically. For example, a tax agency can use a computer algorithm to compare the wage amount reported by an individual and the amount reported by its employer and then automatically notify the individual taxpayers of any discrepancies. Consequently, as individuals earning salaried income are caught evading regardless of whether they are audited, they should not respond to audits (Kleven et al., 2011). On the contrary, there is no comparable automatic cross-checking for VAT enforcement.[1] Thus, tax authorities must rely heavily on audits to discourage VAT evasion (Gomez-Sabaini and Jimenez, 2012; Bergman and Nevarez, 2006).

We collaborated with Uruguay's Internal Revenue Service (from hereon, referred to as IRS) to conduct a natural field experiment with a sample of 20,440 small- and medium-sized firms that are subject to the VAT. For our study, the IRS mailed four different types

[1] While the VAT requires a paper trail, which is a form of third-party reporting, this paper trail is subject to significant limitations. Most important, there is no simple algorithm to detect tax evasion automatically. Second, the paper trail breaks down when reaching the consumer (Naritomi, 2016). Third, firms can also collude with each other to tamper with the paper trail (Pomeranz, 2015).

of letters with information about audits to the owners of each of these firms.[2] Some of the information contained in each of these letters was randomly assigned, with the goal of testing predictions of *A&S*. Using IRS administrative records, we measured the subsequent effects of the information contained in the letters on the firms' compliance with the VAT and other tax responsibilities in the following year. Additionally, we collaborated with the IRS to conduct a post-mailing survey to capture the effect of this information on these firms' subsequent perceptions about audits.

The first part of the experimental design, following the seminal work by Slemrod et al. (2001), measures how informing taxpayers about tax enforcement affects their tax compliance. Firms were randomized into four different letter types: *baseline*, *audit-statistics*, *audit-endogeneity*, and *public-goods*. The *baseline* letter type included brief and generic tax information that the IRS often includes in its communications with firms. The *audit-statistics* letter type was identical to the *baseline* letter, with additional information about the probability of being audited and the penalty rate, based on tax administration statistics. The relevant hypothesis is that adding the *audit-statistics* message to the *baseline* letter will deter tax evasion and thus increase post-treatment tax payments. We also can compare the effects of this *audit-statistics* message with the effects of other types of messages. The *audit-endogeneity* letter provides information about a different feature of the auditing process. The *audit-endogeneity* letter was identical to the *baseline* letter, with an additional message about how evading taxes increases the probability of being audited. The last letter type was designed to provide a benchmark for a message that might increase tax compliance but does not involve the tax audits. The *public-goods* letter was identical to the *baseline* letter, with an additional message describing the social costs from evasion: all the public goods that could be provided if tax evasion was lower.

The first part of the results show that, consistent with Slemrod et al. (2001) and the subsequent literature, informing firms about tax enforcement increases their tax compliance. We find that adding the *audit-statistics* message to the *baseline* letter increases tax payments by about 6.3%. This effect is economically large: the estimated average VAT evasion rate in Uruguay is 26% (Gomez-Sabaini and Jimenez, 2012), meaning that the 6.3% increase equates to a 24% reduction in VAT evasion. This effect also is highly statistically significant and robust to a number of checks, such as alternative specifications and event-study falsification tests. The other message related to audits, *audit-endogeneity*, also significantly increased tax compliance by about 7.4%. This effect is statistically indistinguishable from the 6.3% effect of the *audit-statistics* message and robust. In comparison, the *public-goods* message

[2]Throughout the paper, for simplicity, we refer to firms' perceptions and behavior as a shorthand for firms' owners or managers perceptions and behavior.

had a smaller effect (4.3%) and was statistically insignificant in the baseline specification. Its effect was even smaller in magnitude, and statistically insignificant, in the alternative specifications.

The main goal of this experiment was not to demonstrate that firms react to information about audits but to understand why they react. More precisely, the second and most important part of the experimental design tests the hypothesis that firms react to information about audits as predicted by *A&S*. We provide three tests of *A&S*. The first test exploits survey data on perceptions about audits. If the *audit-statistics* letter increased average compliance, to be consistent with *A&S*, it must be true that this message increased the perceived probability of being audited or the perceived penalty rate. To test this hypothesis, we designed a survey, to be sent months after the firms received the *audit-statistics* and *audit-endogeneity* letters, which measures perceptions about the probability of being audited and the penalty rate.

The second test of *A&S* is based on heterogeneity in the signals provided in the letters. We included exogenous, non-deceptive variation in the information about audit probabilities and penalty rates in the *audit-statistics* letter. To generate this information, we computed the average probabilities and penalty rates using a series of random samples of 50 firms. This sample size was small enough to introduce non-trivial sampling variation in the average probabilities and fines shown to the subjects. That is, a given firm could receive a letter saying that the audit probability is 8%, 10%, or 15%, depending on the sample of similar firms chosen for that particular letter. These random variations in probabilities and penalties shown to the firms allow us to test whether, as predicted by *A&S*, firms evade less when they face higher audit probabilities and higher penalty rates.

As a complement to the *audit-statistics* treatment arm, we designed a separate treatment arm that created exogenous variation in expected audit probabilities in a more direct way. The *audit-threat* letter type was sent to a separate sample of firms that was pre-selected by the IRS for auditing. We randomly divided this set of firms into two groups, one with a 25% probability of being audited and the other with a 50% probability. The *audit-threat* letter informed firms of their audit probability. Again, we can test whether, as predicted by *A&S*, firms evade less when they face a higher audit probability.

The third test of *A&S* exploits heterogeneity by prior beliefs. According to *A&S*, the compliance effect of a given signal about audit probability should depend on the firm's prior belief about that probability. For example, a firm receiving a signal that the audit probability is higher than its prior belief should increase its tax compliance, while a firm receiving a signal lower than its prior belief should decrease its tax compliance. To test this hypothesis, we construct a proxy for prior beliefs about audit probability based on variation in the firms' pre-treatment exposure to audits.[3]

The second part of the results suggests that the effects of the *audit-statistics* letter are not consistent with *A&S*. The results for the first test, based on the survey data, indicate that the *audit-statistics* message reduced the perceived probability of being audited. According to *A&S*, a reduction in the perceived probability of being audited should have reduced tax compliance. On the contrary, we find that the *audit-statistics* message increased average compliance.

The second test shows that, contrary to the *A&S* prediction, the effect of the *audit-statistics* message does change with the signals of audit probability and penalty rates included in the letter. The estimated elasticity of tax compliance, with respect to audit probabilities and penalty rates, is close to zero and precisely estimated. We find qualitatively and quantitatively similar effects between the *audit-statistics* and *audit-threat* treatment arms. Moreover, we compare our experimental estimates to the results from calibrations of *A&S*. We reject the null hypothesis of *A&S*, even under conservative assumptions about how much firms learned from the *audit-statistics* message.

The third test also suggests probability neglect (i.e., that our messages had the same positive effect on tax compliance, regardless of the audit probability communicated in the message or the firm's prior belief about such probability). Contrary to the prediction of *A&S*, the effect of the *audit-statistics* message did not vary with the firm's prior belief about the probability of being audited.

Our findings suggest that small and medium firms may comply with taxes because of the threat of being audited but not necessarily in an optimal manner, as predicted by *A&S*. This leaves open the question of which alternative model best explains the firms' reactions to audits. Models of salience (Chetty et al., 2009) and prospect theory (Kahneman and Tversky, 1979) explain some but not all findings, such as probability neglect. Instead, our preferred interpretation is based on the model of risk-as-feelings (Loewenstein et al., 2001).

The models used for choice under risk are typically cognitive; that is, people make decisions using some type of expectation-based calculus. The risk-as-feelings model proposes that responses to fearsome situations may differ substantially from cognitive evaluations of the same risks. When fear is involved, the responses to risks are quick, automatic, and intuitive and thus neglect the underlying probabilities (Sunstein, 2003; Zeckhauser and Sunstein, 2010). This risk-as-feelings model can explain our finding of probability neglect. Indeed,

[3]Take for instance two firms that have been paying taxes for 10 years and, by chance, one of those firms was audited in the past while the other was not. As a result, the firm that was audited in the past will have a higher belief about the probability of being audited in the future. The implicit assumption is that, due to the scant information on the auditing process, firms may be forming beliefs about the audit probabilities based on their own exposure to audits.

we discuss suggestive evidence that fear plays a significant role in tax compliance. We also discuss policies used by tax agencies around the world that suggest a working knowledge of the risk-as-feelings model.

Our study relates to various strands of literature. First, it belongs to a recent but growing literature that uses field experiments in partnership with tax authorities to study the decisions of individuals to pay taxes. In a seminal contribution, Slemrod et al. (2001) showed that, for a sample of U.S. self-employed individuals, those who were randomly assigned to receive a letter from the Minnesota Department of Revenue with an enforcement message reported higher income in their tax returns. Similar messages about tax enforcement have been shown to have positive effects on tax compliance in other contexts (for recent reviews, see Pomeranz and Vila-Belda, 2018; Slemrod, 2018; Alm, 2019).[4] One standard interpretation in this literature is that taxpayers react to the information about tax enforcement tools and, in line with *A&S*, reduce their evasion to re-optimize their behavior. However, there is no direct evidence in favor or against this interpretation. Our contribution is to fill this gap in the literature.

This paper is closely related to a group of studies testing the predictions of *A&S* in a laboratory setting. For example, Alm et al. (1992) conducted a laboratory experiment in which undergraduate students play a tax evasion game. Subjects can hide income from the experimenter, but some subjects are randomly selected to be audited and, if caught evading, must pay a penalty. The authors show that tax compliance in the game increases significantly with audit and penalty rates, but these effects are economically small and smaller than those predicted by optimizing behavior in the context of *A&S*. The laboratory experiment setting of Alm et al. (1992) and similar studies have several advantages, such as full control over the rules of the game and freedom in the selection of the model parameters. However, these laboratory experiments have two main limitations. First, the subjects are typically undergraduate students playing the tax game for the first time and with no prior experience paying taxes in the real world. In contrast, subjects in our field experiment are experienced firm owners who have been registered with the tax agency, and thus paying taxes, for an average of 15 years. Second, subjects from laboratory experiments typically pay taxes amounting to less than USD 10. In contrast, subjects in our field experiment paid on average USD 11,800 per year, which is in the same order of magnitude as the country's GDP per capita.[5] We contribute to this literature by showing that *A&S* does not fare substantially better in a natural context with experienced subjects and high stakes.

[4]The following are some examples: Slemrod et al. (2001); Kleven et al. (2011); Fellner et al. (2013); Pomeranz (2015); Castro and Scartascini (2015); Dwenger et al. (2016); Perez-Truglia and Troiano (2018).

[5]More specifically, in the 12 months before our experiment firms in our sample paid an average of USD 7,770 in VAT and USD 4,030 in other taxes. In comparison, the GDP per capita in Uruguay was about USD 15,000 in 2015.

Our findings also contribute to the more general debate about the determinants of tax compliance. One of the main puzzles in the literature is that evasion rates seem too low, given the low detection probabilities and penalty rates, especially among smaller firms and self-employed individuals (Luttmer and Singhal, 2014). One traditional explanation for this puzzle is based on tax morale: firms and individuals do not evade taxes because they do not want to (Luttmer and Singhal, 2014). Our evidence suggests an alternative explanation for the puzzle: due to the emotional nature of the decision, taxpayers overreact to the threat of audits. In other words, audits may scare taxpayers into compliance in the same way that scarecrows scare birds. Indeed, this interpretation can explain why, despite the low audit probabilities and penalty rates, taxpayers seem to be significantly concerned about audits: 61% of U.S. taxpayers consider "fear of an audit" to have significant influence on their tax compliance (United States Internal Revenue Service, 2018).[6]

Finally, our paper is also part of a recent but growing literature on behavioral firms (DellaVigna and Gentzkow, 2017). We document two sources of profit-maximizing frictions. First, the fact that firms have biases in beliefs about audit probabilities suggests the presence of information frictions. Second, the fact that tax compliance is inelastic with respect to audit probability and penalty rates suggests the presence of optimization frictions.

The paper is organized as follows. Section 2 discusses the experimental design. Section 3 presents the data sources and discusses the implementation of the field experiment. Section 4 presents the results on the average effect of the *audit-statistics* message, and section 5 presents the different tests of *A&S*. Section 6 discusses the interpretation of the findings. The final section concludes.

2 Experimental Design

Our experiment consisted of a mailing campaign from Uruguay's IRS with multiple treatment arms and sub-treatments. Rather than comparing firms that received a letter to firms that did not, all of our analyses are based on comparisons between firms that received letters with subtle variations in their content. We can thus minimize the potential effects of simply receiving a letter from the tax authority, which might induce compliance on its own as a reminder to pay taxes.

The letters consisted of a single sheet of paper with the name of the recipient in the header, the official letterhead of the IRS, and the scanned signature of the IRS General Director. These letters were folded, sealed in an envelope with the official letterhead of the IRS on the outside, and sent by certified mail, which guarantees direct delivery to the

[6]See section 6 for more details about this survey.

recipient, who must sign upon receipt.

2.1 *Baseline* Letter

The first type of letter is the *baseline* letter, a sample of which is provided in Appendix A.1. The *baseline* letter contained some information about the goals and responsibilities of the tax authority, which the IRS routinely includes in its communications with firms. It explained that the individual was randomly selected to receive this information, that the letter was for information purposes only, and that there was no need to reply or to provide any documentation to the IRS. The letters in the other treatment arms included the same text as in the *baseline* letter, but also included an additional paragraphs, printed in a larger type size and in boldface.

2.2 *Audit-Statistics* Letter

In the *audit-statistics* letter type, we added a paragraph to the *baseline* letter, providing firms with information about the audit and penalty rates. According to the Allingham and Sandmo (1972) model, we expect risk-averse firms to be interested in this information, because it helps them optimize their evasion decisions and potentially increase their bottom line.[7] Furthermore, this information should be particularly valuable in the context of limited information about audits. For instance, it is easy to find information online about factors potentially relevant for firms' decision-making, such as inflation and exchange rates. However, it is extremely difficult to find any information about audit probabilities and actual penalties paid by evading firms. Tax authorities seem to prefer to conceal this information.

Appendix A.2 presents a sample of the *audit-statistics* letter type. The additional paragraph included information about audit probabilities (p) and penalty rates (θ) for a random sample of firms that were similar to the recipient, as follows:

> "On the basis of historical information on similar businesses, there is a probability of *[p%]*that the tax returns you filed for this year will be audited in at least one of the coming three years. If, pursuant to that auditing, it is determined that tax evasion has occurred, you will be required to pay not only the amount previously unpaid, but also a fee of approximately *[θ%]* of that amount."

Note that we communicated the probability that firms will be audited in at least one of the three following years, because IRS experts stated that this was the relevant probability for

[7]We assume that firms in our sample are risk averse, which is plausible since we deal mainly with small and medium firms. However, *A&S* has been generalized to settings with risk-neutral agents (Reinganum and Wilde, 1985; Srinivasan, 1973).

firms' decision-making. Uruguay's tax law indicates that tax audits should cover the previous three years of tax returns and, as a result, the probability that the current year's tax report will be audited is roughly equal to the probability that the firm will be audited at least once over the following three years.

In our sample, the average value of p is 11.7%, and the average value of θ is 30.6%. Tax agencies in most countries do not publish data on the values of p and θ, which makes it difficult to compare the Uruguayan case to other contexts. In the United States, for which some comparable data are available, these two parameters are on the same order of magnitude: self-employed individuals face a p of 11.42% and a base θ of 20%.[8]

The goal of this treatment arm was to generate exogenous variation in the firms' perceptions about audit probabilities and penalty rates. Because of legal and other constraints, we could not assign different firms to different sets of information about these factors. We instead induced non-deceptive, exogenous variation in messages that may affect these perceptions by exploiting the sampling variation in statistics about audits and penalties.

More specifically, we divided the firms into five groups of "similar firms," corresponding to the five quintiles of total VAT payments in the fiscal year before our intervention. For each firm, we then drew a random sample of 50 other firms from the same quintile (i.e., similar firms), from which we computed the averages of p and θ. This randomization strategy generated 940 different combinations of p and θ. These estimates of p and θ were unbiased and consistent with the explanation given in a footnote that we included in the letter, thus information provided to recipients was nondeceptive. The footnote explained how we estimated the values of p and θ:

> "Estimates are based on data from the 2011–2013 period for a group of firms with similar characteristics, for instance, in terms of total revenue. The probability of being audited was calculated as a percentage of audited firms in a random sub-sample of firms. The rate of the fee was estimated as an average of a random sub-sample of audits."

The values of p ranged from 2% to 25%, with an average of about 11.7%. The values of θ ranged from 15% to 68%, with an average of about 30.6%. Figure 1 presents the audit probability and penalty size distribution across five groups by firm size (one in each row)

[8] First, there is an annual probability of being audited of 2.1%, according to the ratio of returns examined for businesses with no income tax credit and with a reported income between USD 25,000 and 200,000 (Table 9a of IRS, 2014). Each audit covers the previous 3 to 6 years, which implies that the the probability that the current year's tax filing will be eventually audited ranges from 5.88% to 11.42%. Second, IRS usually imposes a basic penalty of θ=20%, although the penalties can be higher in more severe cases.

and the distribution of the generated within-group parameters.[9] The vertical line denotes the average probability based on all members of the group. If we based our estimates of p and θ included in the letter on the population of firms, every member of the group would have received the same signal (the vertical line). As we computed p and θ from samples of 50 firms, the sampling variation implies that different members of each group received different signals. For example, Figure 1.a.1. shows that in group 1, the average p for all group members is 8.1%, whereas the histogram depicts the different signals actually sent to firms within the group. These signals center around the average p, but they range from 2.5% to 20%. Note the differences in the vertical lines across groups: for firms in the first quintile, the average randomized p is 8.1%, and this value increases monotonically up to 13.4% for those in the top quintile. This means that a small share of the variation across subjects in the values of p and θ included in the letter results from non-random variation across groups, whereas the within-group variation is fully due to random variation, which is an important factor for the following econometric model.[10]

2.3 *Audit-Threat* Letter

To complement the evidence from the *audit-statistics* sub-treatment, we implemented an alternative way of randomizing perceptions about audit probabilities using an audit-threat letter. We devised a treatment arm that randomly assigned firms to groups with different probabilities of being audited in the following year. A sample of the *audit-threat* letter is presented in A.3. The *audit-threat* letter was identical to the *baseline* letter, with the following additional paragraph:

> "We would like to inform you that the business you represent is one of a group of firms pre-selected for auditing in 2016. A *[X%]* of the firms in that group will then be randomly selected for auditing."

This *audit-threat* treatment arm was applied to a separate experimental sample, a group of high-risk firms selected by the IRS audit department. The recipients of the *audit-threat* letter thus cannot be compared to those of the *baseline* letter. Instead, we randomly assigned the firms in this treatment arm to two groups, one with a probability of being audited in the following year of 25% (X=25%) and another with a probability of being audited twice as large (X=50%). These messages were non-deceptive: the audit department provided a

[9] We constructed the 5 groups according to the quintiles of VAT paid during the tax year before our intervention.

[10] To measure the share of the variation that corresponds to the cluster size, we regress each parameter on the quintiles of VAT payments. Regressing p over pre-treatment VAT quintiles dummies results in $R^2 = 0.118$, while regressing θ over the same dummies results in $R^2 = 0.007$.

commitment that they would conduct audits according to these probabilities in the following year.

2.4 *Audit-Endogeneity* Letter

The *audit-statistics* and *audit-threat* treatment arms conveyed quantitative information about audit probabilities and penalty rates. We also wanted to incorporate into our research design a message about a different aspect of the audit process. Most tax agencies, including Uruguay's, account for firm characteristics when deciding which ones to audit. They assign higher audit probabilities to firms with higher evasion risk. As a result, evading taxes typically increases the probability of being audited. This factor was incorporated as a special case in *A&S*, in which audit probabilities were determined endogenously. If unsuspecting firms learn about the endogenous nature of their audit probabilities, they should revise their tax evasion decisions and reduce the amount of tax evaded.[11]

We used this insight from economic theory to devise the *audit-endogeneity* message about the nature of the audit process. We asked our counterparts at the IRS to use their evasion-risk scores to divide a small sample of firms into two groups: those suspected of evading taxes and those not suspected of evading taxes. We then computed the difference in audit rates from 2011–2013 between the two groups: the rates were approximately twice as high for the former group. We used this information to create the message in the *audit-endogeneity* letter type, which was identical to the *baseline* letter with the addition of the following paragraph (see sample in Appendix A.4):

> "The IRS uses data on thousands of taxpayers to detect firms that may be evading taxes; most of its audits are aimed at those firms. Evading taxes, then, doubles your chances of being audited."

2.5 *Public-Goods* Letter

We also devised a treatment arm to provide a benchmark for the effect of messages intended to increase tax compliance without directly mentioning audits. In line with previous studies (see Blumenthal et al., 2001; Fellner et al., 2013; Pomeranz, 2015; Dwenger et al., 2016), we wanted to include a message that would leverage non-pecuniary motives to increase tax compliance. We designed a non-pecuniary message that was expected to be most effective at increasing compliance by the IRS staff and authorities: a message providing information

[11]Konrad et al. (2016) present suggestive evidence of this mechanism in the context of a laboratory experiment: taxpayers facing a situation where suspicious attitudes toward tax officers increase the probability of being audited increase their tax compliance by 80%.

about the cost of evasion in terms of the provision of public goods, in the spirit of the model of Cowell and Gordon (1988).[12]

The *public-goods* letter is identical to the *baseline* letter, with the addition of a specific paragraph listing a series of services that the government could provide if tax evaders reduced their evasion by 10% (see Appendix A.5 for a sample of the letter):

> "If those who currently evade their tax obligations were to evade 10% less, the additional revenue collected would enable all of the following: to supply 42,000 portable computers to school children; to build 4 high schools, 9 elementary schools, and 2 technical schools; to acquire 80 patrol cars and to hire 500 police officers; to add 87,000 hours of medical attention by doctors at public hospitals; to hire 660 teachers; to build 1,000 public housing units ($50m^2$ per unit). There would be resources left over to reduce the tax burden. The tax behavior of each of us has direct effects on the lives of us all."

We used estimates from different governmental agencies to make the calculations reflected in this message.[13]

2.6 Survey Design

We designed a survey to be conducted with a sample of owners from our main subject pool, months after they received the letters. The IRS, with the support of the Inter-American Center of Tax Administrations and the United Nations, had previously administered a survey on the costs of tax compliance for small- and medium-sized businesses. We collaborated with the tax authority in the design and implementation of a new survey, which included a specific module tailored for our research design. The survey also included seven additional modules, designed by the IRS, about the costs of tax compliance and other topics. Appendix A.6 shows a sample of the email with the invitation to participate in the online survey. We partnered with local and international universities to increase respondent confidence and to highlight that the survey was part of a scientific study and not an audit or compliance exercise by the IRS.

To further ensure trustworthy responses, the IRS assured potential respondents that the survey responses would remain anonymous and that they could not be traced back to specific

[12]This message is also related to the laboratory experiment from Alm et al. (1992), which presents evidence that one of the reasons why people decide to pay taxes is their valuation of the public goods provided by means of the tax revenues.

[13]These agencies were: Administracion Nacional de Educacion Publica (ANEP), CEIBAL, Ministerio de Salud Publica (MSP), Ministerio del Interior (MI), Ministerio de Vivienda, Ordenamiento Territorial y Medio Ambiente (MVOTMA).

individuals or firms. To measure the effect of our experiment on these survey responses, we embedded a code in the survey link to identify the treatment arm of the experiment to which the recipient was assigned. These codes did not uniquely identify any single firm, but they allowed us to link treatment arms and survey responses while maintaining the anonymity of responses.

Appendix A.7 shows a snapshot of our survey module. We designed questions to assess whether the *audit-statistics* message shifted perceptions of the recipients of our letters, by means of the two following questions:

> Perceived Audit Probability: "In your opinion, what is the probability that the tax returns filed by a company like yours will be audited at least in one of the next three years (from 0% to 100%)?"
>
> Perceived Penalty Rate: "Let us imagine that a company like yours is audited and that tax evasion is detected. What, in your opinion, is the penalty (in %) as determined by law that the firm must pay in addition to the originally unpaid amount? For example, a fee of X% means that, for each $100 not paid, the firm would have to pay those original $100 plus $X in penalties."

After each question, we elicited how certain the subject felt about his or her response on a scale from 1 to 4, where 1 is "not at all sure" and 4 is "very sure." For the sake of completeness, we also included a question in the survey to measure the subject's awareness about the endogeneity of audit probabilities. This question and its results are described in detail in Appendix B.3.2.

3 Data Sources and Implementation of the Field Experiment

3.1 Institutional Context

Uruguay is a South American country with an annual GDP per capita of about USD 15,000 in 2015. Total tax revenues (i.e., for all levels of government) were about 19% of GDP in 2015 and, as is common in many countries, VAT represents the largest source of tax revenue, accounting for roughly 50% of the total tax collection.[14] Firms are required to remit VAT

[14]Own calculations based on data from the Central Bank of Uruguay and from the Internal Revenue Service. Other sources of tax revenues include the personal income tax, the corporate tax, and some specific taxes to consumption, businesses and wealth.

payments all along the production and distribution chain.[15] The standard VAT rate is 22%.[16]

Our main focus is on the VAT, which represents the largest tax liability for firms in Uruguay. An important factor in our context is that VAT has limited third-party reporting. The amounts reported for other sources of taxable income are subject to automatic third-party reporting, and thus taxpayers would be caught evading even without audits. For instance, wage income is hard to conceal in modern economies, because employers are usually required to report their employees' earnings to the tax authority. Consequently, tax evasion can be detected and deterred even without conducting an audit (Kleven et al., 2011). Even if VAT requires a paper trail, this reporting has significant limitations in practice. The paper trail breaks down when reaching the consumer, and firms can collude with each other or with customers to tamper with the paper trail, for instance, by offering discounts for sales without receipts (Pomeranz 2015; Naritomi 2016).[17] In this institutional setting and in the case of VAT liabilities for small and medium firms, audits represent the main tool for evasion deterrence. They are the primary mechanism through which tax administrations can detect and deter VAT evasion (Gomez-Sabaini and Jimenez, 2012; Bergman and Nevarez, 2006).

Uruguay does not seem an atypical country in terms of tax evasion and tax morale. According to estimates from Gomez-Sabaini and Jimenez (2012), evasion of VAT in Uruguay was around 26% in 2008. This is the third-lowest rate among the nine Latin American countries included in the study and comparable to evasion rates in more developed economies. For example, similar computations suggest an evasion rate of 22% for Italy in 2006 (Gomez-Sabaini and Moran, 2014).[18] Regarding tax morale, we used data from the 2010–2013 wave of the World Values Survey. According to this data, 77.2% of respondents from Uruguay stated that evading taxes is "Never Justifiable," whereas this proportion is 68.2% on average (population-weighted) for all other Latin American countries and 70.9% for the United States.

[15]Firms may credit VAT paid on input costs (i.e., imports and purchases from their suppliers) against the total sales of goods and services to their costumers (i.e., "tax debit"). They pay VAT to the IRS only on the excess of the total "tax debit" over the tax credit. If the tax credit exceeds the debit, the excess may be carried over for future tax years. While the VAT should in theory be similar in its effects to a retail sales tax, in practice the two types of taxes differ in some substantial aspects (Slemrod, 2008).

[16]A small number of products considered basic necessities either have a 10% rate or are exempt from the tax.

[17]At the time of our experiment, firms the IRS had not yet implemented standardized electronic receipts. In the future, these electronic systems may facilitate and automatize the cross-check of the VAT trail to detect evasion.

[18]Gomez-Sabaini and Jimenez (2012) compute those rates by applying the "indirect" method to estimate tax evasion. This method is based on the comparison of collected VAT to aggregate consumption data from the System of National Accounts (SNA).

3.2 Subject Pool and Randomization

Our experiment was conducted in collaboration with the IRS of Uruguay. As of May 2015, there were 120,142 firms registered in the agency's database. A subsample of 4,597 firms, preselected by the IRS, was put aside for the *audit-threat* sample, which we call the secondary experimental sample. Of the remaining firms, we followed a series of criteria to select our main experimental sample. We first excluded some firms by request of the IRS. For instance, we excluded firms subject to special regimes for VAT payments (very small or very large firms). We also kept in the experimental sample only those firms that made VAT payments in at least three different months in the previous 12-month period and those with a total value of at least USD 1,000.[19]

To maximize the impact of our information provision experiment, we did our best to ensure that the letters would be delivered to the firms' owners.[20] Moreover, in very large firms, the effect of the information could be substantially diluted, as it may not reach the owner or the individuals making decisions about tax compliance. Thus, we excluded from our subject pool firms with a total value exceeding USD 100,000 during the previous 12 months.

These criteria left a subject pool of 20,471 firms for the main experimental sample. All firms were randomly assigned to receive one of the four letter types, with the following distribution: 62.5% were assigned to the main treatment arm (*audit-statistics* letter), and 12.5% were assigned to each of the three remaining letter types (*baseline*, *audit-endogeneity*, and *public-goods*).[21] After removing the roughly 18.5% of letters that were returned by the postal service, the final distribution of letter types was as follows: 10,272 received *audit-statistics*; 2,064 received *baseline*; 2,039 received *audit-endogeneity*; and 2,017 received *public-goods* letters (total N = 16,392). The 4,597 firms in the secondary sample were assigned to receive the *audit-threat* letter. Half were randomly assigned to the message of a 25% audit probability, and the other half to the 50% audit probability. After excluding the 12% of letters returned by the postal service, we were left with 2,015 firms in the 25% probability group and 2,033 firms in the 50% probability group (total N = 4,048). Table 1 provides some descriptive statistics for the firms in our subject pool. The firm characteristics include VAT payments right before we sent the mailing, the age of the firm, the number of employees, and other basic variables. Column (2) corresponds to all firms in the main experimental sample.

[19] The sample selection was conducted in May 2015, so this 12-month period spans from April 2014 to March 2015.

[20] In some cases, owners provide the address of external accountants instead of their own or their firms. We removed from the sample firms that were registered with an accountant's mailing address as their own (the IRS keeps records of addresses for all registered accountants).

[21] The randomization to letter types was stratified by the quintiles of the distribution of VAT payments over the fiscal year before our intervention.

On average, firms had 4.8 employees, had been registered with the IRS for 15.3 years, and 14% had been audited at least once over the previous three years. For comparison, column (1) of Table 1 shows the same statistics for the universe of all registered firms. By design, firms in our experimental sample are smaller, both in terms of number of employees and in level of VAT payments. Finally, column (3) of Table 1 provides statistics about the secondary experimental sample (i.e., *audit-threat* treatment arm). Despite some statistically significant differences in the two groups, the firms are broadly comparable in size. The main difference between firms in the two experimental samples is that the audit rates were 9 percentage points higher in the *audit-threat* sample. This difference is by design, because the IRS selected firms classified as high-risk for this treatment arm, and these firms had a higher propensity to be targeted for audits in the past.

Table 2 allows us to compare the balance of pre-treatment characteristics between firms assigned to the different letter types. Columns (1) through (4) correspond to firms in the main experimental sample. For each characteristic, column (5) presents the p-value of the test of the null hypothesis that the averages for these characteristics are the same across all four letter types. As expected, the differences across letter types are economically small and statistically insignificant. Columns (6) through (8) of Table 2 present a similar balance test, except for the secondary sample used for the *audit-threat* arm. Again, the characteristics are balanced across the two sub-treatments in the *audit-threat* treatment arm.

3.3 Outcomes of Interest

The letters were provided to Uruguay's postal service on August 21, 2015. The vast majority of the letters were delivered during September, and therefore we set August as the last month of the pre-treatment period and October as the first month of the post-treatment period. The main outcome of interest in our study is the total amount of VAT liabilities remitted by taxpayers in the 12 months after receiving the letter.[22] To test for the persistence of our treatment effects, we defined a second period of observation between October 2016 and September 2017 (i.e., up to two years after the intervention).

On average, the total amount of VAT paid by firms that received the baseline letter in the 12-month pre-treatment period was about USD 7,700, whereas the amount for the corresponding post-treatment period was approximately USD 6,500. This negative trend in VAT payments can be explained by the fact that this sample contains a high share of small firms with a high turnover rate. The size of post-treatment VAT payments varied substantially, ranging from the 10th percentile of USD 400 to the 90th percentile of USD

[22]This variable includes all VAT payments, including direct VAT payments and indirect VAT withholdings.

16,550.[23]

We can further break down firms' VAT payments according to their timing. We can observe the date of transfer to the IRS as well as the month for which the payment was imputed. Firms can back-date payments to cover liabilities from previous periods. As firms typically make VAT payments on a monthly basis, they normally cover the current and previous months, which we call concurrent payments. We classified payments for two or more months in the past as retroactive payments. About 99.4% of firms made at least one concurrent payment in the 12-month pre-treatment period, whereas only 23.8% of firms made at least one retroactive payment over this same period.

Finally, although we focused on VAT payments in our analysis, we obtained data from the IRS on the other main taxes paid by the firms, such as corporate income taxes and net worth taxes. These two taxes and the VAT jointly represent more than 96% of the total tax burden of firms. We used payments of these taxes as additional outcomes of interest to study whether firms effectively changed their overall compliance or if they substituted evasion of VAT for that of other taxes.

3.4 Survey Implementation

The IRS communicates mainly by postal mail, and it thus has mailing addresses for all registered firms. It also keeps records of email addresses for a subset of firms that have used their online services. We emailed invitations to all firms in the main experimental sample with a valid email address. Our intention was to survey firms soon after they received our letters, but the survey was postponed due to implementation challenges. We sent email invitations to participate in the online survey to 3,845 firms in May 2016, about nine months after our mailing experiment. Table 1 shows that the firms we invited to the survey (column (5)) were similar in characteristics to those in the main experimental sample (column (2)).

Our purpose was to elicit the beliefs of firm owners. We did not include email addresses that were repeated more than three times in the full sample, as these most likely corresponded to accounting firms representing multiple small and medium enterprises. Even after applying this criteria, the IRS records could not ensure that the registered email address corresponded to the firms' owners. We thus asked the survey respondent to self-identify as one of the following five types: owner, internal accountant, external accountant, manager, or other employee. From the 3,845 recipients that we invited to participate in the survey, we received 2,331 responses (response rate of 60.6%). Of these 2,331, 45% self-identified as owner, 16.2%

[23]Table B.1 in Appendix B.1 presents detailed descriptive statistics about the distribution of pre- and post- treatment payments for firms that received the *baseline* letter type.

as non-owner, and the remaining 38.8% did not provide a response to this question.[24] In our main specification, we excluded responses that identified recipients as non-owners. The results were similar if we included only those who actively identified as owners (reported in Appendix B.3.1). Among respondents who self-identified as owners, the missing data rate for our three key questions was between 19.5% and 23.3%, which is comparable to the average non-response rate for all questions in the survey (17.8%). Finally, per request of the IRS, respondents could skip as many questions as they wanted.

4 Results: Average Effect of Messages

4.1 Econometric Specification

Our main speficication captures the impact of our mailing campaign on our outcomes of interest (subsequent tax payments). These results are obtained by comparing the post-treatment tax payments of firms assigned to the *audit-statistics*, *audit-endogeneity*, and *public-goods* letters with the payments from recipients of the *baseline* letter. We interpret these as the effects of the corresponding messages (*audit-statistics*, *audit-endogeneity*, and *public-goods*) which are the additional information we added to the *baseline* letter.

Consider the sample of firms assigned to either *baseline* letter or one of the other letter types, indexed by j: *audit-statistics*, *audit-endogeneity* or *public-goods*. The main specification is given by the following regression:

$$Y_i = \alpha + \beta \cdot D_i^j + X_i \delta + \epsilon_i \tag{1}$$

The outcome variable (Y_i) is the total outcome in the 12-month post-treatment period (i.e., after the delivery of the treatment letter). D_i^j is a dummy variable that takes the value 0 if i was assigned to the *baseline* letter, and the value 1 if i was assigned to letter type j. Finally, X_i is a vector of control variables. When outcomes are persistent over time, which applies to the case of VAT payments, the use of pre-treatment controls can help reduce the variance of the error term and thus results in gains in statistical power (McKenzie, 2012). Specifically, our main specification includes the outcomes during each of the previous 12 pre-treatment months as control variables (X_i).

This main specification allows us to test our first set of hypotheses: whether providing letters with information about enforcement increases tax compliance. The β coefficient corresponding to the *audit-statistics* message reflects the impact of adding the *audit-statistics*

[24] The non-owner responses are distributed as follows: 4.4% as internal accountant, 5.4% as external accountant, 1.9% as manager, 4.5% as other employee.

message to the *baseline* letter. The findings in the previous literature indicate that this should have a positive effect on tax collection. The β coefficient for the *audit-endogeneity* message, in turn, provides a benchmark for the *audit-statistics* results by capturing the effect of providing other type of information related to audits. Finally, the *public-goods* coefficient is a further benchmark, corresponding to the effect of non-pecuniary incentives.

We rely on a Poisson regression model to estimate our main specification. There are two reasons for the choice of this main specification. First and foremost, the Poisson model allows effects to be proportional – indeed, the coefficients can be readily interpreted as semi-elasticities.[25] Second, the Poisson model naturally accounts for the bunching at zero of the dependent variable. In any case, we present robustness checks with alternative regression models, including OLS and Tobit models. We also estimate separately the effects on the extensive margin. In the interest of transparency, for each coefficient related to post-treatment effects, we also present a falsification test based on pre-treatment "effects"; that is, we re-estimate the model, but with pre-treatment outcomes as the dependent variable. We should expect these pre-treatment effects to be close to zero and statistically insignificant.

4.2 Effect of the *Audit-Statistics* Message

We start by describing the effects of our main treatment, the *audit-statistics* message. We discuss the benchmark results from the other two sub-treatments in the following subsection. Figure 2 summarizes the results from our main specification. In each of the three panels, we plot the difference between the VAT payments of the firms assigned to the *baseline* letter with respect to the payments of firms assigned to each of the other three treatment arms, for the two quarters preceding our mailing campaign and for the eight subsequent quarters.[26] The effects of each of the treatment arms are computed by means of Poisson regressions, so that the coefficients can be directly interpreted as semi-elasticities. These are simple regressions with no additional control variables.

Figure 2.a shows that the *audit-statistics* message had statistically and economically significant effects on VAT payments. For instance, the point coefficient corresponding to the third post-treatment quarter implies that the *audit-statistics* message increased the VAT payments by 8.7% (p = 0.012). The effects are similar in magnitude for all the post-treatment

[25] The Poisson model is based on the following specification: $\log(Y_X) = \alpha + \beta X + \varepsilon$. The effect of a unit change in X can be re-expressed in log-units of the dependent variable, $\beta = \log(Y_{X=x+1}) - \log(Y_{X=x})$. Provided this coefficient is small enough, it can be approximated accurately as a percent-change effect: $\beta = \log(Y_{X=x+1}) - \log(Y_{X=x}) \approx \dfrac{Y_{X=x+1} - Y_{X=x}}{Y_{X=x}}$.

[26] We top-coded all outcomes of interest at the 99.99% percentile to avoid the contamination of the results by outliers.

quarters during the first year (e.g, 7.8%, 5.8%, 8,7% and 5.7% in the first through fourth quarter). During the second year, the effects become smaller and less statistically significant over time. The coefficients on the pre-treatment VAT payments correspond to a falsification test for our experiment. As expected by the random assignment of firms to different treatment arms, the pre-treatment differences in VAT payments are economically small and not statistically significant.

Table 3 presents the baseline regression results. These estimates are obtained by means of Poisson regressions with pre-treatment controls, following the econometric specification described in Section 4.1. The first column presents the main results: the average effects of each of the three treatments (*audit-statistics* in Panel A, *audit-endogeneity* in Panel B, and *public-goods* in Panel C) compared to the outcomes for firms that received the *baseline* letter. The post-treatment coefficients correspond to regressions with total VAT paid in the 12 months after the delivery of the letter as the dependent variable (October 2015 - September 2016). Additionally, as falsification tests, the pre-treatment coefficients correspond to regressions with total VAT paid in the 12 months prior the delivery of the letter as the dependent variable (September 2014 - August 2015).

The post-treatment coefficient of *audit-statistics* (first column, panel a.) indicates that firms receiving this message paid 6.3% more VAT in the 12 months after the intervention on average. This effect is not only statistically significant (p = 0.013), it is also economically substantial. Using the estimated average evasion rate of 26% from Gomez-Sabaini and Jimenez (2012), the effect amounts to a reduction in the evasion rate of 24% ($= \frac{6.3\%}{26\%}$). As expected, the "effect" on pre-treatment outcomes (our falsification test) is close to zero (-0.8%), statistically insignificant, and even more precisely estimated than the corresponding post-treatment effect.[27]

In terms of previous findings in the literature, the effects of our *audit-statistics* treatment are not directly comparable to those of the audit message from Pomeranz (2015) because the messages differed in content, and because the two studies cover firms from different countries and with different characteristics. Nevertheless, Table 4 from Pomeranz (2015) indicates that the deterrence letter in that study led to an increase in VAT payments of 7.6%, which is similar in magnitude and statistically indistinguishable from the 6.3% effect of our *audit-statistics* message. Moreover, our results are consistent with a broader literature that finds effects of messages about enforcement on tax compliance in a variety of contexts: self-employed income in the United States (Slemrod et al., 2001), wage income taxes in Denmark (Kleven et al., 2011), individual public-TV fees in Austria (Fellner et al., 2013), individual municipal taxes

[27]The standard error on the pre-treatment coefficient is 0.021, which is smaller than the corresponding standard error of 0.025 for the post-treatment coefficient.

in Argentina (Castro and Scartascini, 2015), an individual church tax in Germany (Dwenger et al., 2016), and U.S. tax delinquencies in the United States (Perez-Truglia and Troiano, 2018).

The second column in Table 3 replicates the analysis for the second year after the treatment, i.e. October 2016 - September 2017. For this period, the effect of the treatment is less than half than that of the first year, and it is not statistically significant, even though the precision of the estimate is similar in the two time periods. This is consistent with the pattern of effects by quarter depicted in Figure 2.a, which shows that the effects in the second year fall substantially when compared to those of the first year. These results are not unexpected. Over time, individuals may forget the information conveyed by the letter, or it may become less salient. Individuals may also update their beliefs and perceptions for other reasons, for instance because of new events such as audits and information campaigns. This is consistent with previous evidence on the effects of tax enforcement messages. Figure 2 in Pomeranz (2015) shows that the effects of the main information campaign in that study were also substantially higher in the first 12 months after their intervention, and fell substantially and became virtually zero in the 18th month (the last plotted result in the Figure).

Table 3 also presents results for complementary outcomes. As discussed in the previous section, firms in Uruguay make payments for their current liabilities, but also for taxes that correspond to previous periods — because they owe past taxes, because they revise their accounts and correct past mistakes, or because they impute invoices that were not available at the time of the original payment. When firms that engage in tax evasion face a heightened threat of being audited, we can expect them to increase their tax payments (reduce their evasion) in the future, but we can also expect them to retroactively revise their payments for previous time periods to reduce or eliminate their past evasion. We explore this possibility with the results presented in columns (3) and (4) of Panel a in Table 3, which split the effects of our treatment arms on retroactive and concurrent payments. These results correspond to the first year after the treatment. The messages about audits had an economically and statistically significant effect on retroactive payments. Indeed, because of the much lower baseline rate, the effects on retroactive payments are larger in magnitude than the effects on the concurrent payments. For instance, the effect for *audit-statistics* message is 38.1% ($p = 0.004$) for the retroactive payments and 4.4% ($p = 0.087$) for the concurrent payments.

We have so far established that firms in the *audit-statistics* treatment arms increased their VAT payments compared to recipients of the *baseline* letter. Our analysis focuses on VAT liabilities, which represents the largest fraction of tax payments by firms in our sample/ However, our letters referred to taxes in general and did not mention VAT nor any other specific tax. Given the presence of other tax liabilities, the effects we reported on VAT may not represent a net increase in tax payments: firms may increase their evasion (i.e., reduce their payments) of other taxes they are liable for, crowding out payments or substituting evasion. The results in columns (5) and (6) of Table 3 shed light on these issues. Column (5) presents the effect on all other taxes paid (mostly the corporate income tax): the effects on payments of other taxes are as economically and statistically as significant as those on VAT payments: the *audit-statistics* had an effect of 7.7% (p = 0.038) on other tax payments. Column (6) shows that the results are robust if we look at the effects on the sum of VAT and other taxes: the *audit-statistics* message increased this outcome by a statistically significant 5.2%.

4.3 Benchmarks to the *Audit-Statistics* Message

Our research design allows us to compare the effects of the *audit-statistics* message to the effects of two alternative messages, one related to audits (*audit-endogeneity*) and another not related to audits nor other forms of enforcement (*public-goods*).

Figure 2.b shows that the *audit-endogeneity* treatment arm also induced a significant change in VAT payments compared to the *baseline* letter. The quarterly effect of adding the *audit-endogeneity* message is positive and statistically significant in the first post-treatment year, and ranges between 10% and 12.4%. These effects are slightly higher than those of the *audit-statistics* treatment arm, but these differences are not statistically significant at standard levels. As in the case of the *audit-statistics* message, the effects of the *audit-endogeneity* treatment are persistent over the first year, but diminish over the second post-treatment year. The results of the falsification test are also similar to the ones observed in the *audit-statistics* treatment: the differences for the two pre-treatment quarters are economically small and not statistically significant for the *audit-endogeneity* treatment arm.

This overall pattern of results is confirmed by the regression results presented in Panel b in Table 3. The coefficient in the first column indicates that the *audit-endogeneity* message increased subsequent VAT payments by 7.4% (p = 0.021), with no "effect" on pre-treatment outcomes, as expected. The effect during the first post-treatment year is similar in magnitude to the 6.3% effect of our *audit-statistics* message (the difference is not statistically significant at standard levels). The effects over the second post-treatment year are also consistent with those of the *audit-statistics* message: the coefficients for the second year are smaller in magnitude than those for the first year, and they are not statistically significant. The *audit-endogeneity* message also had an economically and statistically significant effect on both retroactive (31.4%) and concurrent payments (6.1%), as in the case of the *audit-statistics* message. Finally, the results in columns (5) and (6) of Table 3 indicate that, again as in the case of the *audit-statistics* message, the *audit-endogeneity* message did not crowd out other tax payments: it increased non-VAT tax payments by 8.4%.

We now turn to the effects of the *public-goods* treatment arm. Figure 2.c indicates that adding this message to the *baseline* letter did not affect VAT payments as consistently as the two other treatments. While there are differences in post-treatment VAT payments between firms assigned to the *public-goods* letters and those assigned to the *baseline* letter, inspection of coefficients over time reveals that the differences in the post-treatment period were similar to the differences in the pre-treatment period.

Given these pre-treatment differences, the assessment of the effect of the *public-goods* message requires us to control for them. This is presented in Panel c in Table 3. The effect for the first year of the post-treatment period as a whole for the *public-goods* message is positive at 4.3%, and falls to 0.6% in the second year, and neither of these two coefficients are statistically significant at standard levels (p-values of 0.147 and 0.879 respectively). Compared to the other two treatment arms, these effects are smaller, but due to the lack of precision they are statistically indistinguishable. The coefficients are not statistically significant in any of the alternative specifications. The *public-goods* message did not have a statistically significant effect either on retroactive or concurrent VAT payments (p-values of 0.304 and 0.879, respectively). The effect of the *public-goods* message on other tax payments (column 5 in Panel c in Table 3) is close to zero (0.1%) and statistically insignificant. The effect of the *public-goods* message on total tax payments is also small (1.8%) and statistically insignificant. Due to the precision of our estimates, we cannot rule out that the public-goods message may have had some effect on tax payments. However, this findings is consistent with the robust finding that moral suasion messages do not have significant effects on tax evasion in a variety of contexts (Blumenthal et al., 2001; Fellner et al., 2013; Castro and Scartascini, 2015; Dwenger et al., 2016; Meiselman, 2018; Perez-Truglia and Troiano, 2018).[28]

Appendix B.2.1 presents some additional robustness checks, including regression results based on alternative specifications such as OLS, Tobit and Probit models, as well as focusing on the extensive margin, and using an alternative independent source of administrative data. We show that the main findings are qualitatively and quantitatively similar to those presented in this section, and thus robust to these alternatives.

[28]There are some exceptions in the literature, such as the effect of displaying norms on the timing of tax payments (Hallsworth et al., 2017).

5 Tests of *A&S*

The results presented in the previous section are broadly consistent with the evidence in the previous literature: providing information about audits significantly increased tax compliance. In this section, we present additional evidence to establish whether the effects of the *audit-statistics* treatment are driven by the *A&S* mechanism. The subsections below present three different tests of *A&S*.

5.1 First Test of *A&S*: Effects on Perceptions

According to *A&S*, the *audit-statistics* message should have a positive effect on tax compliance if it increased the perceived probability of being audited, the perceived value of the evasion fine, or both. We explore this hypothesis based on data from our post-treatment survey. This survey data consists of 365 firms in the *audit-statistics* group and 137 in what we refer to as the pooled control group with individuals who did not receive information related to audits.[29]

Figure 3.a and 3.b depict the distributions of perceptions about audit probabilities and penalty rates, respectively, as elicited from the survey. The shallow bars with solid borders correspond to the perceptions of firms that received the *audit-statistics* message. The shaded gray bars depict the distribution of perceptions for individuals from firms in the pooled control group. The red dashed curve, in turn, corresponds to the distribution of signals sent to firms in the *audit-statistics* letters. The comparison between the shaded bars and the red curve from Figure 3.a suggests that, on average, respondents in the control group substantially overestimated the probability of being audited. While our administrative data on audits indicates a probability of about 11.7%, the mean perception for the control group is 40.7% (p-value < 0.01 for the difference). This finding of overestimation of audit probabilities is consistent with prior survey evidence (Harris and Associates 1988; Erard and Feinstein 1994; Scholz and Pinney 1995).[30]

Moreover, survey participants reported being confident on their responses even though their estimates were substantially off: only 16.2% of those in the control group reported being

[29]The survey sample size was substantially smaller than that of our experimental sample. To increase the statistical power of our test, we defined this control group by pooling subjects from the *baseline* and the *public-goods* groups, since both received messages with no specific information about audit probabilities or fines. Appendix B.3.1 shows that the results are similar, but less precisely estimated, when we only use recipients of the *baseline* letter for the control group.

[30]However, the prior survey evidence was based on responses from wage-earners, for whom the misperception of audit probabilities is mostly inconsequential due to widespread third-party reporting (Kleven et al., 2011). On the contrary, the financial stakes of misperceiving audit probabilities can be substantial in our context.

"Not sure at all" about their perceived probability of audit (on a four point scale, ranging from "Not sure at all" to "Very confident").[31] Indeed, even for the subgroup of individuals from the control group who reported to be "Very confident" about their guesses, their average belief was, if anything, slightly more biased: 42.1%, still substantially higher than the actual probability of 11.7%.

Conversely, the comparison between the shaded bars and the red curve from Figure 3.b suggests that the average belief about the penalty rates was unbiased: the actual average penalty computed from administrative data for the experimental sample is 30.7%, while the mean perceived penalty is 30.5% in the control group.

The positive bias in the perceived audit probability can be explained by the availability heuristic bias (Kahneman and Tversky, 1974). According to this model, individuals judge the probability of an event by how easily they recall instances of it. Even though audits are rare, the fact that they may be visible for colleagues and even sometimes salient in the media may induce firms to perceive them to be more frequent than they actually are. Indeed, there is evidence that individuals overestimate the probabilities of a wide range of rare events of a similar nature (Lichtenstein et al., 1978; Kahneman et al., 1982).

The systematic upward bias in perceptions of audit probabilities and the results presented in the previous section, showing that the *audit-statistics* message increased tax compliance, are not consistent with *A&S*. Since our letters provided information about audit probabilities that was lower than firms' beliefs, the *A&S* framework would predict a reduction tax compliance. Our post-treatment survey allows us to test this hypothesis more directly. From the assignment of firms to the different treatment groups, we can measure the effect of the *audit-statistics* letter on the perceived p and θ. The shallow bars with solid borders in Figures 3.a and 3.b depict the distribution of perceptions for respondents from firms in the *audit-statistics* treatment arm. Inspection of Figure 3.a indicates that, compared to respondents in the treatment group, recipients of the *audit-statistics* message reported on average a lower perceived probability of being audited, from an average of 40.7% in the pooled control group to an average of 35.2% in the *audit-statistics* group (p-value of the difference 0.03). Meanwhile, Figure 3.b shows that the *audit-statistics* message had a small effect on the perceived penalty rate, decreasing it from an average of 30.5% for the pooled control group to an average of 29.9% for the *audit-statistics* group, with this difference being statistically insignificant (p-value of 0.85).

This analysis only provides a lower bound for the effect of our letters on perceptions. While we are confident that our certified letters reached firms' owners, we cannot be as confident about whether the owner was the same person who received the email invitation

[31] A similar share (18.1%) reported to be "Not sure at all" about their guess for the penalty rate.

to complete the survey. Moreover, the survey responses took place over nine months after our letters were sent, so it is feasible that recipients had forgotten the mailing's messages at that point, or that they had acquired additional information in the meantime. Thus, the true effect of the audit-statistics on the average subjective audit probability was probably even more negative than what we report above.

To sum up, the evidence suggests that the *audit-statistics* treatment arm may have reduced the average perceived audit probability. This evidence, and the positive effects of this treatment on tax compliance, are jointly inconsistent with *A&S*, according to which this message, by reducing perceived audit probabilities, should have reduced tax payments. Consistent with this evidence, Appendix B.3.2 suggests that the effect of the *audit-endogeneity* message is not due to an update of recipients' beliefs about the degree of endogeneity of audit probabilities, because recipients were already well aware of this endogeneity.

One important caveat for this test and those we present in the following subsections is that our analysis relates to the behavior of the average firm. While the average firm does not behave as *A&S* predicts, we cannot rule out that some of the firms behave in a manner consistent with *A&S*. In other words, it is possible that some firms updated their perceived probability upwards because of the information contained in the letter and increased their tax payments as a consequence.

Another caveat with this test is that it is based on survey data, which has some limitations. However, we can address some of those challenges. A first concern is that our subjects may be confused about the meaning of an "audit." However, our survey data provides direct evidence on the contrary. Among the 145 responses from the pooled control group, 10.3% of firms reported that they were audited in the past three years. Since this share (10.3%) is close to the actual share of firms that were audited (11.7%), it seems that respondents understood the definition of an audit correctly.

A second concern with survey data is that subjects may have difficulties responding about percentages and probabilities. However, this may be a less of a concern in our subject pool, which is comprised of business owners who should be familiar with fractions and probabilities. At the very least, they need some rudimentary arithmetic and understanding of percentages to compute the VAT and other tax liabilities.

A last source of concern with our survey data is that, in some circumstances, respondents may report probabilities of exactly 50% as a way of expressing their uncertainty (Bruin et al., 2002; Bruine de Bruin and Carman, 2012). Responses of exactly 50% are somewhat common in our data: among individuals in the pooled control group, 33.3% of responses about the perceived audit probability and 16.4% of responses about the penalty rate are exactly equal to 50%. To assess the extent of this concern, we follow the standard method from Bruin et al. (2002) and Bruine de Bruin and Carman (2012). We measure certainty using a 1-4 scale from "Not sure at all" (1) to "Very confident" (4). If there were large differences in certainty between individuals who responded exactly 50% and the rest, that would cast doubts on the validity of the survey responses. On the contrary, we find modest differences in certainty between the two groups.[32] As a complementary robustness check, we can conduct our analysis ignoring 50% responses and treating them as missing data. Even with this conservative approach, individuals in the control group still substantially over-estimate the probability of being audited: these respondents report an average perception of 31.5%, compared to the actual probability of 11.7%.

In the following sections, we provide alternative tests of *A&S* that do not rely on survey data.

5.2 Second Test of *A&S*: Heterogeneity with Respect to Signals

5.2.1 Reduced-Form Results

The second test is based on the differential effects of the values of the signals provided in the letters. According to *A&S*, the effects of the audit-statistics message should be increasing in the signals about the audit probability (p) and the penalty rate (θ). The random variation we introduced in the p and θ conveyed in our *audit-statistics* letters allow us to test this hypothesis directly. Figure 4 starts with a less parametric look at the data. This figure uses the main specification from before,[33] looking effect of the *audit-statistics* message on VAT payments, but broken down by decile of the signals included in the letter. Figure 4.a presents the effect of the *audit-statistics* message by decile of the signal of p included in the letter. In the *A&S* framework, we should expect that very low signals of p should reduce tax compliance (since they most likely reduce the firms' perceived probability of audits), whereas the effect should become larger, and turn positive at some point, as we increase the value of the signal about p.

The coefficients plotted in Figure 4.a, however, indicate that the effect of the *audit-statistics* letter is not related to the value of p included in the letter. The coefficients are similar in magnitude for the whole range of values from $p = 2\%$ all the way up to $p = 25\%$. For example, the effect is 8.8% and statistically significant (p-value of 0.016) for the lowest

[32] In the pooled control group, the average certainty for perceived audit probability is 2.18 for individuals who responded a value of exactly 50%, and 2.56 for individuals who responded a different value (p-value of difference 0.09). For the responses about the perceived penalty rate, the average certainty is 2.18 for individuals who responded 50%, and 2.48 for those who responded another value (p-value of difference 0.35).

[33] Additionally, we control for dummies for the quintiles of pre-treatment VAT payments, from which we drew the sample to calculate p_i and θ_i.

decile ($p \in [2\%, 5\%]$), and it is 5.6% and marginally significant (p-value of 0.065) for firms in the upper decile ($p \in [19\%, 25\%]$). Moreover, the resulting slope (in dashed red line), while positive as predicted by *A&S*, is economically small (0.0014) and statistically insignificant (p-value=0.637).

Figure 4.b provides an analysis similar to that of Figure 4.a, but instead of looking at the heterogeneity by audit probabilities (p) it shows the heterogeniety by penalty rates (θ). According to *A&S*, we should expect a positive relationship between the effect of the *audits-statistic* letter and the value of θ included in the letter. Figure 4.b shows evidence to the contrary: the coefficients are similar for the whole range of values from $\theta = 15\%$ to $\theta = 68\%$, with the slope being negative, economically small and statistically significant.

One potential confounding factor for our findings about the lack of effect of variations in p and θ is that some subjects might have interpreted the *audit-statistics* message *per se* as a signal that their firms were under the IRS radar, above and beyond the factual information conveyed in the message. We were careful to mitigate this concern in the design of our mailings. For instance, we highlighted the fact that the letter recipients were randomly selected. Nevertheless, some individuals may have ignored or overlooked this cue. However, even if the recipients learned something from the receipt of the audit-statistics message, there is no reason why they should not learn about the content of the message as well. In other words, the test presented above continues to be valid, as *A&S* would still predict that the *audit-statistics* message should have a differential effect depending on the values of p and θ.

To address this concern more directly, we use the *audit-threat* treatment arm, in which the tax agency made an explicit threat to every recipient and thus is not subject to this concern. Figure 5 depicts the difference in the evolution of VAT payments over time between the two sub-treatments in the *audit-threat* arm, corresponding to audit probabilities of 50% and 25%. We find no systematic difference between the two groups in post-treatment VAT payments. This complementary test further reinforces the result that, contrary to the prediction of *A&S*, tax compliance does not depend on the exact probability being audited.

5.2.2 Elasticities with respect to p and θ

For a more direct test of A&S, we can quantify the effects of the *audit-statistics* and *audit-threat* sub-treatments in a way that can be contrasted to the quantitative predictions of *A&S*. For the *audit-statistics* treatment arm, we use the following model:

$$Y_i = \alpha + \gamma_p \cdot p_i + \gamma_\theta \cdot \theta_i + \sum_{g=2}^{5} \pi_g \cdot I_{\{i \in g\}} + X_i \delta + \epsilon_i \tag{2}$$

where $p_i \in (0,1)$ is the signal about the audit probability included in the letter sent to subject i, and $\theta_i \in (0,1)$ is the signal about the penalty rate included in the letter sent to the same firm. The $I_{\{i \in g\}}$ variables correspond to a set of dummies for quintiles of pre-treatment VAT payments, which are the groups from which we drew the sample of "similar firms" to calculate p_i and θ_i. Including these controls ensures that we only exploit the exogenous variation in p_i and θ_i induced by our experimental design – that is, the heterogeneity due to sampling variation. Since we are using a Poisson regression model, γ_p and γ_θ can be directly interpreted as elasticities. For instance, an estimate of $\gamma_p = 1$ would imply that a 1 percentage point increase in the audit probability conveyed in the letters increased VAT payments by 1%. From *A&S*, we can expect that $\gamma_p > 0$ and $\gamma_\theta > 0$ – i.e., firms tax payments are increasing in the perceived probability of audit and evasion penalty rates. Moreover, we can compare the values of these regression estimates to the predictions from calibrations of *A&S*.

A similar econometric model can be used for firms assigned to the *audit-threat* letter:

$$Y_i = \alpha + \gamma_p \cdot p_i + X_i\delta + \epsilon_i \tag{3}$$

where $p_i \in \{0.25, 0.50\}$ is the audit probability included in the *audit-threat* letter sent to firm i. As in the previous regression, *A&S* implies a positive estimate of γ_p.

Panel (a) in Table 4 presents the results from the econometric model of equation 2. Column (1) of Table 4 presents estimates of the elasticities of VAT payments with respect to the values of p and θ conveyed in the *audit-statistics* sub-treatments. The elasticity with respect to the audit probability in the first year after the treatment is 0.030 (SE of 0.236, p-value of 0.897). This means that increasing p by 1 percentage point would increase VAT payments by a mere 0.03%. The elasticity with respect to the penalty rate is −0.118 (SE 0.115, p-value of 0.304), which implies that increasing θ by 1 percentage point would reduce VAT payments by 0.118%. The estimates are close to zero, not statistically significant at standard levels and precisely estimated. The precision implies that we can rule out even moderate elasticities: the 90% confidence interval for the audit probability excludes elasticities above 0.418, and the 90% confidence interval for the penalty rate excludes elasticities above 0.071.

It should be noted that the pre-treatment falsification test does not yield any statistically significant effect, and that the results are similar (no statistically significant elasticities) for the other specifications: for the second year (column (2)), by timing of the payment (columns (3) and (4)), and by type of tax (columns (5) and (6)). In most cases, the estimates are not statistically significant. There are some exceptions, such as the coefficient associated with the audit probability for retroactive payments, but in this case the differences were statistically significant before the treatment too, and therefore this result is probably spurious.

As complementary evidence, panel (b) of Table 4 presents the results of the *audit-threat* treatment arm in regression form, following the econometric model of equation 3. While the *audit-threat* messages implies an elasticity with respect to p of 0.376, this estimate only borderline significant at the 10% level (p-value of 0.073) and economically small. Moreover, the pre-treatment (falsification) coefficient (-0.342) is also borderline statistically significant at the 10% level (p-value of 0.055), indicating that the small post-treatment effect may be spurious.

Appendix B.2.3 present a series of robustness checks (alternative specifications based on OLS, Tobit and Probit models, and using an alternative data source for the dependent variable), and the results are similar. An additional robustness check, presented in the same Appendix, shows that results are robust if, instead of estimating the elasticities with respect to p and θ separately, we estimate the elasticity with respect to $p * \theta$ (i.e., the expected penalty per dollar evaded).

Taken together, this evidence suggests that firms did not react to the values of the parameters conveyed in our letters. Since we devoted a large fraction of our subject pool to this treatment arm, these elasticities are quite precisely estimated.

Moreover, we can assess the economic magnitude of these differences by benchmarking these elasticities with the quantitative predictions of *A&S*. The exact predictions of *A&S* depend on the specific model at hand and on the values of the underlying parameters. We provide calibrations under a number of settings that have been considered in the literature, allowing for social preferences (Luttmer and Singhal, 2014), endogenous audit probabilities (Allingham and Sandmo, 1972; Yitzhaki, 1987), misperceptions about audit parameters (Alm et al., 1992) and non-audit detection (Acemoglu and Jackson, 2017). We calibrate different combinations of these models to match the average VAT evasion rate in Uruguay (26%, as estimated by Gomez-Sabaini and Jimenez, 2012), and compute the elasticity of tax payments with respect to p and θ under each of the alternative calibrations.

The details and results from these calibrations are presented in Appendix C. Since all the different calibrations lead to elasticities in the same order of magnitude, the results are similar regardless of the calibration that we use. In our preferred calibration, we find an elasticity of tax payments with respect to the audit probability of 4.55, and an elasticity of tax payments with respect to the penalty rate of 3.48. We can test the null hypothesis that the elasticities with respect to p and θ in the main specification of the *audit-statistics* presented in column (1) of Table 4 are equal to those in our preferred *A&S* calibration. We can reject the null that the elasticity is 4.55 for the audit probability, and that it is 3.48 for the penalty rate (both tests with p-values<0.001). These calibrated elasticities are also well beyond the confidence intervals for the coefficients estimated for the *audit-threat* model.

It should be noted that our discussion is based on the implicit assumption that a letter conveying the message of a 1 percentage point higher signal of p or θ will increase the perception of the parameter by the recipient by 1 percentage point. This is probably a strong assumption: some individuals may not have read the letter in its entirety, they may have not entirely believed in the content of our message, or may not have updated their prior by the full value of the signal. For a benchmark, we can compare our setting with studies of learning from economic variables, such as the inflation rate (Cavallo et al., 2017), the cost of living (Bottan and Perez-Truglia, 2017), and housing prices (Fuster et al., 2018). These studies find that for each percentage point increase in the feedback given to subjects, the average individual updates beliefs about half a percentage point. If we assume this rate of learning, then we should double the elasticities estimated in our regressions before comparing them to the calibrations of *A&S*. Under this assumption, we can still reject the null that the estimated and the calibrated elasticities of tax compliance with respect to the audit probability are equal (p-value<0.001 for each γ_p and γ_θ).

Since we have survey data on the *audit-statistics* group, we can estimate the learning rate directly. However, this exercise faces two significant challenges. First, we elicited these beliefs with a survey conducted nine months after the information was provided, so the effect of the information may have decayed substantially. For example, Cavallo et al. (2017), shows that the effect of information on beliefs decays by about half in a matter of just three months (similar findings are reported in Bottan and Perez-Truglia, 2017 and Fuster et al., 2018). The second limitation is the small sample size (365 observations). With those caveats in mind, the survey data suggests that a percentage point increase in the signal about the audit probability provided in the letter increased the perceived audit probability nine months later by 0.397 (SE 0.288) percentage points. This point estimate must be taken with a grain of salt because it is imprecisely estimated and thus statistically insignificant at conventional levels. However, it is reasurring that the point estimate suggests a learning rate that is consistent with the learning rates from other studies used as a benchmark above.

Moreover, we can reproduce the analysis under an extremely conservative assumption about the magnitude of the learning rate. Even if we assumed that for each percentage point difference in the letter individuals only adjusted their beliefs by one tenth of a percentage point, we would still fail to reject the null hypothesis that the estimated elasticities are equal to those in the *A&S* calibration.

Our estimates are similar to those obtained in laboratory experiments. For example, Alm et al. (1992) find an elasticity with respect to the audit probability of 0.169 (comparable to our estimate of 0.030), and with respect to the penalty rate of 0.037 (comparable to our estimate of -0.118). Indeed, the elasticities reported in Alm et al. (1992) are statistically indistinguishable from the elasticities reported in our study.

The findings from the audit-statistics and audit-threat arms are also consistent with some results from other field experiments. Dwenger et al. (2016) conducted a field experiment in the context of a local church tax in Germany for which enforcement was very lax. In one of their treatment arms, they included information with different probabilities of audits ($p = 0.1$, $p = 0.2$, or $p = 0.5$). They do not compute elasticities with respect to p that we can compare with ours. However, consistent with our probability neglect result, the effect on increased compliance and increased payments is very similar (and statistically indistinguishable) for the three treatment arms. Another related experiment, Kleven et al. (2011), included a treatment arm with two different audit probabilities. Consistent with our results, they find economically negligible differences in tax compliance between individuals assigned to different audit probabilities.[34] However, the evidence from Kleven et al. (2011) is not inconsistent with *A&S* because, unlike in our context, their subjects face automatic third party reporting.[35]

5.3 Third Test of *A&S*: Heterogeneity with Respect to Prior Beliefs

5.3.1 Measuring Prior Beliefs

In the *A&S* framework, firms with different prior beliefs about the probability of being audited should react differently to signals and information about this probability. To test this hypothesis, we need a measure for prior beliefs for a particular firm. We construct a proxy of prior beliefs based on the firm's own audit history.

The intuition behind this approach is that, since there is little publicly available information about audit probabilities, firms may form their beliefs based on their own audit experience. For instance, when a firm registers with the tax authority, its initial belief may follow the beta distribution with parameters $\{\alpha_0, \beta_0\}$. Assume that firm i has been registered for T_i years before our mailing campaign, and during this period it has experienced $N_i \leq T_i$ audits. If firm i is Bayesian, its belief about annual probability of being audited should follow a beta distribution with parameters $\{\alpha_1 = \alpha_0 + N_i,\ \beta_1 = \beta_0 + T_i - N_i\}$. The mean of that belief should be $\frac{\alpha_0+N_i}{\alpha_0+\beta_0+T_i}$. In turn, this implies a belief of the probability of being audited at

[34] In one of their treatments, they send letters to individuals stating large audit probabilities of $p = 50\%$ and $p = 100\%$. Compared with a group that did not receive any letter, they find that the letters had a positive and significant effect on declared income and tax liability. The differential effects between these two conditions, while is statistically significant, is economically negligible: an increase in the signal about probability of audit from 50% to 100% increases reported income by 0.025% and taxes paid by 0.05%.

[35] They conduct their experiments with wage earners, for whom evasion is automatically detected through third-party reporting without the need for audits. As a result, *A&S* predicts that, consistent with their evidence, wage earners should report their wage incomes truthfully regardless of the probability of being audited (Kleven et al., 2011).

least once in the following three years of $\hat{p}_i = 1 - \left(1 - \frac{\alpha_0+N_i}{\alpha_0+\beta_0+T_i}\right)^3$. In our main specification, we generate these proxies setting $\{\alpha_0 = 0.13, \beta_0 = 1\}$. This baseline calibration generates an average belief that matches the actual average probability in our administrative data, and we offer robustness tests using alternative calibrations.[36]

We can use the survey data to validate this proxy for the prior belief. Among the 145 responses from the pooled control group, 10.3% of firms reported that they were audited in the past three years and the remaining 89.7% reported that they were not. We find that respondents from firms that were audited in the recent past reported a higher average perceived probability of being audited (63.9%) than individuals from firms that were not recently audited (38.1%). This different is not only economically large, but also statistically significant (p-value<0.001).[37] This evidence suggests that, consistent with our proxy, firms are using their own audit history to form beliefs about the probability of being audited in the future.

5.3.2 Results

In the *A&S* framework, the effect of the *audit-statistics* letter on tax compliance should be larger for firms with relatively low priors for the audit probability ($\hat{p}$) compared to those with relatively high values of $\hat{p}$. More specifically, the *audit-statistics* conveyed, on average, a signal of $\hat{p} = 11.7\%$. The effect of this message on compliance should thus have been positive for the group with $\hat{p} < 11.7\%$: i.e., on average, the signal should increase their perceived audit probability. On the contrary, the effect of the *audit-statistics* message should be negative for firms with $\hat{p}$>11.7%: on average, the signal should reduce their perceive probability of being audited.

Figure 6 presents the results. Figure 6.a presents a binned scatterplot with the treatment effect of the *audit-statistics* letter for the four quartiles of $\hat{p}$. This figure includes a vertical dashed line at 11.7%, the average message about the audit probability conveyed by our letters. In contrast to the predictions from *A&S*, we fail to find a negative relationship between the effect of the audit-statistics message and the value of the prior belief: the slope is negative (-0.0213), but economically small and statistically insignificant (p-value of 0.146).

Figure 6.b presents a more direct test, which combines heterogeneity in prior beliefs with

[36]Since we only have information about audits for firms in our sample for the previous 15 years, we set the maximum firm age at 15 to compute these priors.

[37]As reported in a follow-up paper (Bergolo et al., 2018), the indicator of recent audits is the single most important predictor of perceived audit probabilities among a host of different factors. The recent audit experience also has a positive effect on the perceived penalty rate, but this effect is less significant: respondents from firms that were audited recently report an average perceived penalty rate of 40.0%, compared to 29.4% for respondents from firms that were not audited recently, although this difference is statistically insignificant (p-value of 0.201).

heterogeneity in signals. Instead of grouping firms by their prior beliefs, we group then by the difference between their prior and the specific signal sent to each firm in its personalized letter. The intuition is that the difference between the prior and the signal is the "surprise" conveyed by our information treatment. We included a vertical dashed line at 0, i.e., at the point were firms receive signals equal to their priors. The effect of the *audit-statistics* letter on compliance should be decreasing in $\hat{p} - p_{signal}$, positive for the group with $\hat{p} - p_{signal} < 0$ (i.e., those for whom the signal was higher than their prior) and negative for the group with $\hat{p} - p_{signal} > 0$ (i.e., those receiving a signal indicating that they were overestimating the audit probability). The results in Figure 6.b are consistent with the results from Figure 6.a: the slope of the relationship (in dashed red line) is negative (-0.0129) but economically small and statistically insignificant (p-value of 0.515).

Appendix B.2.2 shows that these results are robust to a number of checks. For instance, the results are similar if we calibrate $\hat{p}$ to be centered at the average perceived audit probability instead of the actual audit probability. Moreover, the results are consistent with a less parametric version of the test, in which we simply compare the effects of the *audit-statistics* message between firms that were audited in the past and firms that were not.

6 Discussion

Our experiment was designed to test the null hypotheses of *A&S*. Given that our evidence rejected these null hypotheses, it remains unclear which alternative model is best suited to explain the findings. We consider some of these alternative models in this section. To facilitate the discussion, we organize the main findings in three groups:

- *Increased compliance*: on average, the *audit-statistics* message had a positive effect on tax compliance.
- *Reduced subjective probability*: on average, the *audit-statistics* message decreased the perceived probability of being audited.
- *Probability neglect*: the effect of the *audit-statistics* message did not depend on the audit probability included in the letter nor on the firm's prior belief about this probability.

Below we discuss the plausibility of different models:

***A&S* with Salience.** One natural model to consider is that of salience (Chetty et al., 2009). According to this model, firms behave as if the probability of detection and the penalty rate are zero, unless these parameters are made salient to them. This model could reconcile the findings of increased compliance and reduced subjective probability: even if those who were sent the messages adjusted their perceived audit probabilities downwards, they would have behaved as if those probabilities were zero if they had not received those messages. However, the salience model fails to fit other features of our findings. First, by definition, salience models imply short-lived effects. A reminder about a non-salient tax should affect the behavior of an agent only when receiving the information, but not later on. This prediction contradicts our evidence about the persistent increased compliance from firms that received our *audit-statistics* letter. This persistence remained for months after the messages were transmitted. Salience models also are inconsistent with our finding of *probability neglect* (i.e., making salient a high probability of audit should have a stronger impact than information about a low probability of audit).

***A&S* with Prospect Theory.** It is possible to incorporate prospect theory (Kahneman and Tversky, 1979) into *A&S* (e.g., Dhami and al Nowaihi, 2007). This extension of the model, however, is unlikely to explain our findings. Most important, prospect theory is unlikely to explain our finding of *probability neglect*. Although differences between extremely low probabilities can be ignored under prospect theory, the range of probabilities in our context was far from what is normally considered extremely low (e.g., in the *audit-threat* arm, the probabilities were 25% versus 50%).

Risk-as-feelings. Our preferred interpretation is based on the model of risk-as-feelings (Loewenstein et al., 2001). The models used for choice under uncertainty are typically cognitive, in the sense that agents make decisions using some type of expectation-based calculus. The risk-as-feelings model proposes that responses to fearsome situations may differ substantially from cognitive evaluations of the same risks (Loewenstein and Lerner, 2003).[38] When fear is involved, the responses to risks are quick, automatic, and intuitive, and they tend to neglect the cost-benefit calculus. A key prediction of this model is that feelings about risk will be mostly insensitive to changes in probability, which is known in the literature as probability neglect (Sunstein (2002); Zeckhauser and Sunstein, 2010), or the fear that makes individuals focus on the downside of outcomes and thus ignore the underlying likelihoods. There is evidence of probability neglect in a range of fearsome situations involving electric shocks, arsenic, abandoned hazardous waste dumps, pesticides, and anthrax (Sunstein (2003); Zeckhauser and Sunstein, 2010).

[38]A related concept, the affect heuristic, corresponds to the quick, automatic and intuitive evaluations of risky situations based in emotions, which might be used as shortcut for more complex evaluations of risk (Slovic et al., 2004). Borrowing Kahneman (2003) terminology for the dual system model of the human mind, emotions might influence the intuitive system.

This model of risk-as-feelings can reconcile all of our key findings. It explains the finding of *probability neglect*, as well as the findings of *increased compliance* and *reduced subjective probability*. Even if the perceived probability of an audit decreased among the treatment subjects, they still may be scared into paying more taxes, because they did not rely on cognitive evaluations of probabilities anyway. Our interpretation suggests that taxpayers overreact to the threat of audits. In other words, audits may scare taxpayers into compliance in the same way that scarecrows scare birds. Indeed, beyond tax compliance, fear has been found to drive excessive caution in other contexts (Loewenstein et al., 2001).[39]

There is some evidence that individuals may have an emotional reaction when dealing with tax audits and, more generally, the tax authority. A survey by the United States Internal Revenue Service (2018) indicates that 61% of U.S. taxpayers consider "fear of an audit" to have significant influence on their tax compliance decisions.[40] This is also reflected in the popular press. For example, a The Washington Post (2016) article claims that "a lot of people are super scared of the Internal Revenue Service" and that its powers "can instill a lot of fear." The fear of the tax authority can be extreme enough to be considered a phobia (New York Times, 2009). Laboratory experiments also have confirmed the role of fear in tax compliance. Coricelli et al. (2010) conducted a fairly typical tax evasion game in the laboratory and measured how emotional arousal affected tax evasion decisions. They showed that the intensity of emotional arousal predicts whether individuals evade and by how much. In a related laboratory setting, Dulleck et al. (2016) showed a significant correlation between tax compliance and physiological markers of stress during the tax reporting decision.

In other areas of public policy, the risk-as-feelings heuristic can be a problem, because it distorts facts and promotes irrational judgment, leading to suboptimal decisions from a pure risk-assessment perspective. For example, Zeckhauser and Sunstein (2010) and others discuss cases involving regulation of nuclear power, vaccines, and other emotion-arousing issues.

For tax collection, these emotional biases might have positive implications, at least from the tax authority's perspective. In fact, some anecdotal evidence suggests that tax authorities use the risk-as-feelings model to foster tax compliance. In the United States, for example, a disproportionately large number of tax enforcement press releases covering criminal convictions and civil injunctions are released during the weeks immediately preceding Tax Day, presumably to scare taxpayers into preparing compliant returns (Morse, 2009; Blank and

[39]For example, fear to terrorist attacks can make people choose other, more dangerous forms of transport; and fear of shark attacks can lead to unnecessary legislation (Sunstein, 2002, 2003; Zeckhauser and Sunstein, 2010).

[40]More precisely, 32% of respondents claim that "fear of audits" exert "a great deal of an influence" and 29% "somewhat of an influence" in whether they honestly report and pay their taxes. In comparison, audits are perceived to be as strong of a deterrent as third-party reporting: 66% of respondents stated that "third party reporting (e.g., wages, interest, dividends)" influence their tax compliance.

Levin, 2010). Moreover, some tax experts claim that the IRS "likes [targeting] celebrities because they get the most bang for their buck in terms of publicity" to "scare the public into complying" (Forbes, 2008).

Furthermore, the risk-as-feelings framework indicates that vivid imagery can effectively instill fear and biases in risk evaluations (Slovic et al., 2004; Zeckhauser and Sunstein, 2010). Coincidentally, tax agencies seem to resort to vivid images in some of their advertising campaigns. For example, in the United Kingdom, the tax agency posts banners in all sorts of public spaces. One of those banners, reproduced in Figure 7, shows a pair of eyes peeking threateningly through a gash in the paper. The poster reads, "If you've declared all your income you have nothing to fear." The tax authority in Formosa, Argentina, was perhaps even less subtle: its ad campaign starred a monster kidnapping tax evaders and individuals in arrears with their taxes.[41] The tax authority claims that the use of the monster was supposed to be humorous, but they also may have been trying to send a subliminal message (Clarin, 2016). Similarly, a TV advertisement in the United States showed the IRS as "something like poltergeist coming out of a TV set and the world falling apart," followed by the phrase, "Have you filed your income tax?" (United Press International, 1988).

These campaigns and anecdotes suggest that some tax administrations may be leveraging fear for tax collection. Even if administrations leverage existing fear rather than instill it, relying on such tactics is not obvious from a normative perspective, since the ethics of such campaigns could be questionable. Moreover, actively promoting fear could have unintended negative effects, such as imposing negative psychological stress on taxpayers. [42]

7 Conclusions

The canonical model of Allingham and Sandmo (1972) predicts that firms evade taxes by optimally trading off between the costs and benefits of evasion, but it is unclear whether real-world firms react to audits in this way. We designed a large-scale field experiment in collaboration with Uruguay's tax authority to assess the factors behind firms' evasion behavior and their reactions to audits. Our findings indicate that firms do increase their tax compliance when informed about the auditing process. However, we do not find this reaction to be consistent with the predictions of *A&S*. For example, the information about audits decreased (rather than increased) the perceived probability of being audited; also, the effects of our messages about audit probabilities were independent of the signal we conveyed and

[41] They used an aboriginal mythical monster, the "Pombero," a sort of boogeyman that according to folklore took away misbehaving children.

[42] For a discussion on the ethical and practical issues with the use of communication efforts to increase tax compliance, see for example Morse (2009).

of the firms' prior beliefs. Models of salience are consistent with the increased compliance we observed and with the reduced perceived perception of audit probabilities, but they are not consistent with our findings of probability neglect. We argue that all three findings can be reconciled by the risk-as-feelings model, which highlights the role of emotions in decision-making and which predicts that agents might exhibit probability neglect in dreaded or feared situations, like paying taxes.

We conclude by discussing some policy implications. In the traditional framework of *A&S*, the relevant policy lever is the number of audits: the tax agency must find the point at which the marginal cost of an additional audit equals the expected marginal benefit (i.e., higher tax revenues). Our findings suggest that small and medium firms face significant information and optimization frictions when reacting to audits. These frictions introduce new levers for policy-making. For example, tax agencies can decide whether to be transparent about the auditing process,[43] whether to contact taxpayers to remind them of the auditing process, and whether to make the costs of being a tax cheat salient and vivid through advertisement campaigns.[44] Indeed, we discussed anecdotal evidence that some tax agencies may already have a working knowledge of these new policy levers. For example, some tax agencies seem to avoid transparency about the auditing process while increasing visibility of enforcement actions around tax day. Some even refer to fear in their advertisement campaigns. However, there is no direct evidence on whether these policies effectively increase tax compliance or whether they have unintended effects, such as instigating enough fear in taxpayers to make them so anxious and unhappy that it trumps the positive effects of tax revenues. As stated by Alm (2019), in a recent review of the literature, "the role of emotions in tax compliance decisions remains largely unexamined." Our results highlight the need for more research on probability neglect in the decision to pay taxes. Moreover, additional research should examine the role of emotions on other important economic choices beyond tax compliance.

[43]On the one hand, our evidence indicates that increasing transparency about the audit probability would reduce the average perceived probability of being audited, which could in turn reduce tax compliance. On the other hand, our finding of probability neglect suggests that, in the end, the reduction in perceived audit probability may not affect tax compliance.

[44]For a practical discussion on how to implement this type of policy, including the drawbacks, see Morse (2009). Furthermore, this same principle can be used to improve compliance with other laws. For instance, Dur and Vollaard (2019) show experimental evidence that salience of tax enforcement can be used to reduce illegal garbage disposal.

References

Acemoglu, D. and M. O. Jackson (2017). Social Norms and the Enforcement of Laws. *Journal of the European Economic Association 15*(2), 245–295.

Allingham, M. G. and A. Sandmo (1972). Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics 1*, 323–338.

Alm, J. (2019). What motivates tax compliance? *Journal of Economic Surveys 33*(2), 353–388.

Alm, J., B. Jackson, and M. McKee (1992). Estimating the Determinants of Taxpayer Compliance with Experimental Data. *National Tax Journal 45*(1), 107–114.

Alm, J., G. H. McClelland, and W. Schulze (1992). Why do people pay taxes? *Journal of Public Economics 48*(1), 21–38.

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy 76*(2), 169–217.

Bergman, M. and A. Nevarez (2006). Do Audits Enhance Compliance? An Empirical Assessment of VAT Enforcement. *National Tax Journal 59*(4), 817–832.

Bergolo, M., R. Ceni, G. Cruces, M. Giaccobasso, and R. Perez-Truglia (2018). Misperceptions about Tax Audits. *AEA Papers and Proceedings 108*, 83–87.

Blank, J. D. and D. Z. Levin (2010). When Is Tax Enforcement Publicized? *Virginia Tax Review 30*.

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), 125–138.

Bottan, N. L. and R. Perez-Truglia (2017). Choosing Your Pond: Location Choices and Relative Income. *NBER Working Paper* (23615).

Bruin, W. J. A. B. d., P. S. Fischbeck, N. A. Stiber, and B. Fischhoff (2002). What number is "fifty-fifty"? Redistributing excess 50% responses in risk perception studies. *Risk Analysis 22*(4), 725–735.

Bruine de Bruin, W. and K. G. Carman (2012). Measuring Risk Perceptions: What Does the Excessive Use of 50% Mean? *Medical Decision Making 32*(2), 232–236.

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

Cavallo, A., G. Cruces, and R. Perez-Truglia (2017, July). Inflation Expectations, Learning, and Supermarket Prices: Evidence from Survey Experiments. *American Economic Journal: Macroeconomics 9*(3), 1–35.

Chetty, R., A. Looney, and K. Kroft (2009). Salience and Taxation: Theory and Evidence. *The American Economic Review 99*(4), 1145–77.

Clarin (2016, September). Polemica campana en formosa al que no paga los impuestos se lo lleva el pombero. *Clarin*.

Coricelli, G., M. Joffily, C. Montmarquette, and M. C. Villeval (2010). Cheating, Emotions, and Rationality: An Experiment on Tax Evasion. *Experimental Economics 13*(2), 226–247.

Cowell, F. A. and J. P. F. Gordon (1988). Unwillingness to pay: Tax evasion and public good provision. *Journal of Public Economics 36*(3), 305–321.

DellaVigna, S. and M. Gentzkow (2017). Uniform Pricing in US Retail Chains. Working Paper 23996, National Bureau of Economic Research.

Dhami, S. and A. al Nowaihi (2007). Prospect theory versus expected utility theory: Why Do People Pay Taxes? *Journal of Economic Behavior and Organization 64*(1), 171–192.

Dulleck, U., J. Fooken, C. Newton, A. Ristl, M. Schaffner, and B. Torgler (2016). Tax compliance and psychic costs: Behavioral experimental evidence using a physiological marker. *Journal of Public Economics 134*, 9 – 18.

Dur, R. and B. Vollaard (2019). Salience of law enforcement a field experiment. *Journal of Environmental Economics and Management 93*(C), 208–220.

Dwenger, N., H. Kleven, I. Rasul, and J. Rincke (2016). Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany. *American Economic Journal: Economic Policy 8*(3), 203–232.

Erard, B. and J. S. Feinstein (1994). The Role of Moral Sentiment and Audit Perceptions in Tax Compliance. *Public Finance 49*, 70–89.

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), 634–660.

Forbes (2008, July). Pity The Celebrity Taxpayer. *Forbes*.

Fuster, A., R. Perez-Truglia, and B. Zafar (2018). Expectations with Endogenous Information Acquisition: An Experimental Investigation. *NBER Working Paper*.

Gomez-Sabaini, J. C. and J. P. Jimenez (2012). Tax structure and tax evasion in Latin America. *Macroeconomics of Development Series 118.*

Gomez-Sabaini, J. C. and D. Moran (2014). Tax policy in Latin America Assessment and guidelines for a second generation of reforms. *Macroeconomics of Development Series 133.*

Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics 148*, 14–31.

Harris, L. and I. Associates (1988). 1987 taxpayer opinion survey. *Washington, DC: Internal Revenue Service Document*.

Kahneman, D. (2003). A Perspective on Judgement and Choice: Mapping Bounded Rationality. *American Psychologist 58*(9), 697–720.

Kahneman, D., P. Slovic, and A. Tversky (Eds.) (1982). *Judgment under uncertainty: heuristics and biases.* Cambridge ; New York: Cambridge University Press.

Kahneman, D. and A. Tversky (1974). Judgment under Uncertainty: Heuristics and Biases. *Science 185*(4157), 1124–1131.

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica 47*(2), 263–291.

Kleven, H. J., M. B. Knudsen, T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or Unable to Cheat? Evidence from a Randomized Tax Audit Experiment in Denmark. *Econometrica 79*(3), 651–692.

Konrad, K. A., T. Lohse, and S. Qari (2016). Compliance With Endogenous Audit Probabilities. *Scandinavian Journal of Economics*.

Lichtenstein, S., P. Slovic, B. Fischho, M. Layman, and B. Combs (1978). Judged frequency of lethal events. *Journal of experimental psychology: Human learning and memory 4*(6), 551.

Loewenstein, G. and S. Lerner (2003). The role of affect in decision making. In R. Davidson, K. Scherer, and H. Goldsmith (Eds.), *Handbook of Affictive Sciences.* Oxford: Oxford University Press.

Loewenstein, G. F., E. U. Weber, C. K. Hsee, and N. Welch (2001). Risk as feelings. *Psychological Bulletin 127*(2), 267–286.

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

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics 99*(2), 210–221.

Meiselman, B. S. (2018). Ghostbusting in detroit: Evidence on nonfilers from a controlled field experiment. *Journal of Public Economics 158*, 180 – 193.

Morse, S. C. (2009). Using Salience and Influence to Narrow the Tax Gap. *Loyola University Chicago Law Journal 40*, 483.

Naritomi, J. (2016). Consumers as Tax Auditors.

New York Times (2009, April). A Paralyzing Fear of Filing Taxes. *New York Times*.

Perez-Truglia, R. and U. Troiano (2018). Shaming tax delinquents. *Journal of Public Economics 167*, 120–137.

Pomeranz, D. (2015). No Taxation Without Information: Deterrence and Self-Enforcement in the Value Added Tax. *The American Economic Review 105*(8), 2539–2569.

Pomeranz, D. and J. Vila-Belda (2018). Taking State-Capacity Research to the Field: Insights from Collaborations with Tax Authorities.

Reinganum, J. F. and L. L. Wilde (1985). Income tax compliance in a principal agent framework. *Journal of Public Economics 26*(1), 1–18.

Scholz, J. T. and N. Pinney (1995). Duty, Fear, and Tax Compliance: The heuristic basis of citizenship behavior. *American Journal of Political Science 39*, 2.

Slemrod, J. (2008). Does It Matter Who Writes the Check to the Government? The Economics of Tax Remittance. *National Tax Journal 61*.

Slemrod, J. (2018). Tax Compliance and Enforcement. *Journal of Eonomic Literature Forthcoming*.

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), 455–483.

Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor (2004). Risk as analysis and risk as feelings: some thoughts about affect, reason, risk, and rationality. *Risk Analysis: An Official Publication of the Society for Risk Analysis 24*(2), 311–322.

Srinivasan, T. N. (1973). Tax Evasion: A Model. *Journal of Public Economics 2*(4), 339–346.

Sunstein, C. (2002). Probability Neglect: Emotions, Worst Cases, and Law. *Yale Law Journal 112*(1), 61–107.

Sunstein, C. R. (2003). Terrorism and probability neglect. *Journal of Risk and Uncertainty 26*(2-3), 121–136.

The Washington Post (2016, August). That is NOT the IRS Calling You! *The Washington Post*.

United Press International (1988). Psychologist takes issue with irs scare tactic. *UPI-United Press International*.

United States Internal Revenue Service (2018). Comprehensive Taxpayer Attitude Survey (CTAS) 2017 Executive Report. Publication 5296 (Rev. 3-2018) Catalog Number 71353Y, Department of Treasury, Washington, D.C.

Yitzhaki, S. (1987). On the Excess Burden of Tax Evasion. *Public Finance Review 15*(2), 123–137.

Zeckhauser, R. and C. R. Sunstein (2010). Dreadful Possibilities, Neglected Probabilities. In E. Michel-Kerjan and P. Slovic (Eds.), *The Irrational Economist: Making Decisions in a Dangerous World*, pp. 116–123. New York: Public Affairs Press.

Table 1: Comparison of Firm Characteristics for Different Groups: All Firms, Firms in Main Sample, Firms in the Secondary Sample and Firms Invited to the Online Survey

	All firms (1)	Experimental Sample: Main (2)	Experimental Sample: Secondary (3)	Invited to the survey (4)
Share paid VAT taxes (3 months pre-mailing)	0.78	0.93	0.89	0.93
	(0.42)	(0.26)	(0.31)	(0.26)
Amount of VAT paid (3 months pre-mailing)	3.72	1.89	1.74	1.89
	(11.55)	(2.83)	(4.25)	(2.98)
Years registered in tax agency	14.21	15.26	19.44	14.46
	(14.85)	(17.16)	(12.84)	(10.08)
Share audited between 2013-2015	0.06	0.10	0.14	0.08
	(0.32)	(0.40)	(0.46)	(0.35)
Number of employees	12.65	4.84	4.89	6.43
	(302.97)	(26.60)	(5.76)	(53.77)
Share retail trade sector	0.13	0.22	0.33	0.15
	(0.34)	(0.41)	(0.47)	(0.36)
Share Agricultural, forest and others	0.03	0.03	0.03	0.02
	(0.17)	(0.17)	(0.18)	(0.15)
Share construction sector	0.03	0.03	0.03	0.03
	(0.17)	(0.17)	(0.17)	(0.18)
Share other sector	0.84	0.73	0.61	0.80
	(0.37)	(0.45)	(0.49)	(0.40)
N	120,142	16,392	4,048	3,845

Notes: Average characteristics in different subsamples of the universe of firms registered in the tax agency (standard deviations in parentheses). Column (1) includes all firms that submitted at least one payment in 2014 or 2015. Column (2) includes the subset of firms selected for the experimental sample according to the criteria described in section 3.2. Column (3) represents a group of high risk firms that were selected from a special sample defined by the IRS and received the *audit-threat* letter. Column (4) corresponds to firms with valid e-mail addresses on file with the IRS, and therefore selected to participate in the on-line survey conducted after the experiment. All data is based on administrative tax records (monthly payments, annual tax returns and auditing registers). Robust standard errors in parentheses.

Table 2: Balance of Firm Characteristics across Treatment Groups

	Main Sample					Secondary Sample		
	Audit Statistics (1)	Public Goods (2)	Audit Endogeneity (3)	Baseline (4)	p-value test (5)	Audit Threat (25%) (6)	Audit Threat (50%) (7)	p-value test (8)
Share paid VAT taxes (3 months pre-mailing)	0.92	0.94	0.93	0.93	0.18	0.90	0.89	0.54
	(0.00)	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	
Amount of VAT paid (3 months pre-mailing)	1.87	1.96	1.93	1.91	0.56	1.74	1.75	0.95
	(0.03)	(0.07)	(0.07)	(0.06)		(0.10)	(0.09)	
Years registered in tax agency	15.34	14.75	15.70	15.01	0.27	19.45	19.42	0.94
	(0.17)	(0.22)	(0.54)	(0.22)		(0.28)	(0.29)	
Share audited between 2013-2015	0.11	0.10	0.09	0.10	0.30	0.13	0.15	0.38
	(0.00)	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	
Number of employees	4.81	4.66	4.88	5.09	0.96	4.83	4.88	0.80
	(0.26)	(0.54)	(0.57)	(0.64)		(0.13)	(0.12)	
Share retail trade sector	0.22	0.22	0.21	0.23	0.78	0.33	0.32	0.40
	(0.00)	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	
Share Agricultural, forest and others	0.03	0.03	0.03	0.03	0.77	0.03	0.04	0.12
	(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	
Share construction sector	0.03	0.03	0.03	0.03	0.85	0.03	0.03	0.33
	(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	
Share other sector	0.73	0.73	0.73	0.72	0.95	0.61	0.62	0.54
	(0.00)	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	
Observations	10,272	2,017	2,039	2,064		2,015	2,033	

Notes: Averages for different pre-treatment firm-level characteristics, by treatment group and type of sample (robust standard errors in parentheses). The main sample includes all firms selected as described in section 3.2. The secondary sample includes high risk firms selected by the IRS. Standard errors in parentheses. The last column of each sample reports the p-value of a test in which the null hypothesis is that the mean is equal for all the treatment groups. Data on VAT amount and firm characteristics comes from administrative tax records (including monthly payments, annual tax returns and auditing registers). Robust standard errors in parentheses.

Table 3: Average Effects of *Audit-Statistics*, *Audit-Endogeneity* and *Public-Goods* Messages on VAT and Other Tax Payments by Time Horizon and Payment Timing

	By Time Horizon		By Payment Timing		By Tax Type	
	First Year (1)	Second Year (2)	Retroactive (3)	Concurrent (4)	Non - VAT (5)	VAT + Non-VAT (6)
a. Audit - Statitstics (N= 10,272) vs Baseline (N= 2,064)						
Post-Treatment	0.063**	0.032	0.381***	0.044*	0.077**	0.052**
	(0.025)	(0.031)	(0.131)	(0.026)	(0.037)	(0.026)
Pre-Treatment	-0.008	-0.003	-0.132	-0.005	0.016	0.037
	(0.021)	(0.025)	(0.105)	(0.022)	(0.045)	(0.026)
b. Audit - Endogeneity (N= 2,039) vs Baseline (N= 2,064)						
Post-Treatment	0.074**	0.038	0.314**	0.061*	0.084**	0.079***
	(0.032)	(0.039)	(0.139)	(0.033)	(0.041)	(0.031)
Pre-Treatment	-0.006	0.072**	0.008	-0.003	0.006	0.027
	(0.027)	(0.030)	(0.125)	(0.027)	(0.056)	(0.031)
c. Public - Goods (N= 2,017) vs Baseline (N= 2,064)						
Post-Treatment	0.043	0.006	0.195	0.032	0.001	0.018
	(0.030)	(0.036)	(0.134)	(0.031)	(0.048)	(0.029)
Pre-Treatment	-0.004	0.031	-0.192	0.002	-0.047	-0.004
	(0.026)	(0.028)	(0.124)	(0.026)	(0.046)	(0.027)

Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Robust standard errors in parentheses. The results are based on Poisson regressions, so coefficients can be interpreted directly as semi-elasticities. Panel a. compares the *audit-statistics* message with the *baseline* letter, while results in panels b. and c. replicate the comparison for *audit-endogeneity* and *public-goods* messages. In the first row of each panel, the dependent variable is the amount of VAT payments (in dollars) after receiving the letter. The second row presents a falsification test in which we estimate the same regression but using the amount contributed before receiving the mailing (pre-treatment) as the dependent variable. All regressions are estimated with a set of monthly controls that correspond to the year before the outcome: for the the post-treatment outcome regression, we include monthly VAT payment controls from September, 2014 to August, 2015, and for the pre-treatment outcome we include the same variables for the September 2013 – August 2014 period. We also restrict the analysis to firms that effectively received the letter as reported by the postal service. Columns (1) and (2) report the effect of treatment by time horizon. Column (1) presents the estimates of the effect for the first year (October 2015 – September 2016) while Column (2) reports the effect of the letter in the second year after the treatment (October 2016 – September 2017). Columns (3) and (4) present the first year effect of treatment on retroactive (3) and concurrent (4) VAT payments. Columns (5) and (6) report the first year results by type of tax. Column (5) presents the effect of the treatment on other (non-VAT) tax payments, while column (6) reports the effect on the total amount of taxes paid by the firms during the same period.

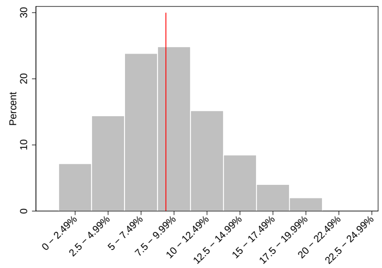
Table 4: Elasticities of Tax Payments with Respect to Audit Probability and Penalty Rate, *Audit-Statistics* and *Audit-Threat* Sub-Treatments

	By Time Horizon		By Payment Timing		By Tax Type	
	First Year (1)	Second Year (2)	Retroactive (3)	Concurrent (4)	Non - VAT (5)	VAT + Non-VAT (6)
a. Audit - Statitstics (N= 10,272)						
Audit Probability (%)						
Post-Treatment	0.030	0.170	-2.136**	0.162	0.127	0.094
	(0.236)	(0.236)	(1.040)	(0.246)	(0.440)	(0.266)
Pre-Treatment	0.025	-0.041	-1.984**	0.123	0.192	0.111
	(0.115)	(0.069)	(0.806)	(0.121)	(0.388)	(0.169)
Penalty Size (%)						
Post-Treatment	-0.118	-0.233*	1.044	-0.191*	-0.291	-0.179
	(0.115)	(0.122)	(0.742)	(0.112)	(0.180)	(0.112)
Pre-Treatment	-0.001	0.049	0.032	-0.001	-0.376**	-0.136
	(0.088)	(0.034)	(0.470)	(0.093)	(0.168)	(0.086)
b. Audit - Threat Letters (N= 4,048)						
Audit Probability (%)						
Post-Treatment	0.376*	0.378*	0.831	0.335	0.164	0.295*
	(0.210)	(0.220)	(0.944)	(0.216)	(0.185)	(0.170)
Pre-Treatment	-0.342*	-0.214	0.013	-0.289	-0.219	-0.308**
	(0.178)	(0.162)	(0.647)	(0.182)	(0.163)	(0.142)

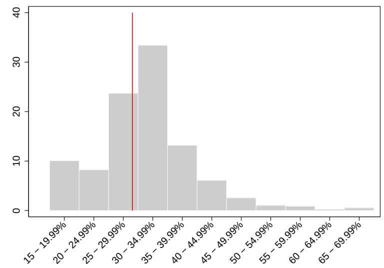
Notes: * significant at the 10% level, ** at the 5% level, *** at the 1% level. Robust standard errors in parentheses. Panel a. presents the effect of providing different information regarding p and θ in the *audit-statistics* message. Panel b. compares the two *audit-threat* messages, i.e. the 50% threat of audit vs. the 25% threat of audit. Rows (1) and (3) of panel a. present the effect of an additional percentage point of p and θ (respectively) in the information included in the letters on post-treatment VAT payments. The results are based on Poisson regressions, so coefficients can be interpreted directly as semi-elasticities. Regressions are estimated using monthly pre-treatment controls and the stratification variable used to randomize the parameters. Rows (2) and (4) present a falsification test in which we estimate the same regression using pre-treatment information as the dependent variable. All regressions are estimated with a set of monthly controls that correspond to the year before the outcome: for the the post-treatment outcome regression, we include monthly VAT payment controls from September, 2014 to August, 2015, and for the pre-treatment outcome we include the same variables for the September 2013 – August 2014 period. We also restrict the analysis to firms that effectively received the letter as reported by the postal service. Row (1) in panel b. presents the post-treatment effect of receiving the *audit-threat* letter with a probability of 50% relative to receiving the letter with the 25% probability treatment. Row (2) of panel b. replicates the estimates for the pre-treatment outcomes. Columns (1) and (2) report the effect of treatment by time horizon. Column (1) presents the estimates of the effect for the first year (October 2015 – September 2016) while Column (2) reports the effect of the letter in the second year after the treatment (October 2016 – September 2017). Columns (3) and (4) present the first year effect of treatment on retroactive (3) and concurrent (4) VAT payments. Columns (5) and (6) report the first year results by type of tax. Column (5) presents the effect of the treatment on other (non-VAT) tax payments, while column (6) reports the effect on the total amount of taxes paid by the firms during the same period.

Figure 1: Distribution of Statistics Shown in *Audit-Statistics* Letters by VAT Payment Quintiles

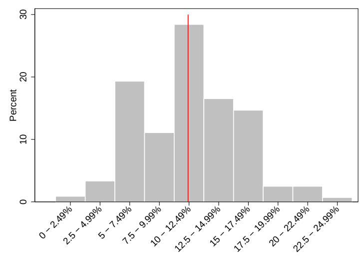
a.1. p: Group 1



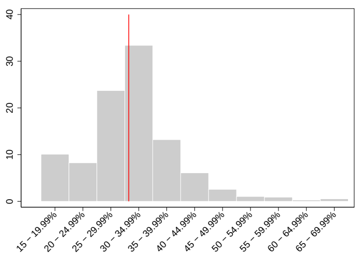
b.1. θ: Group 1



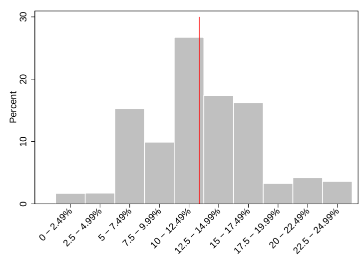
a.2. p: Group 2



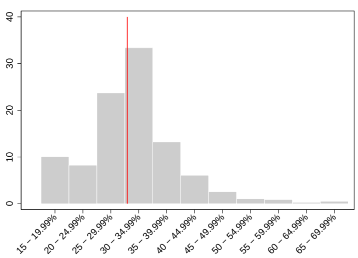
b.2. θ: Group 2



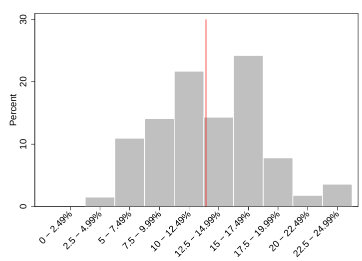
a.3. p: Group 3



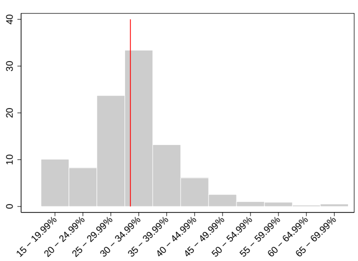
b.3. θ: Group 3



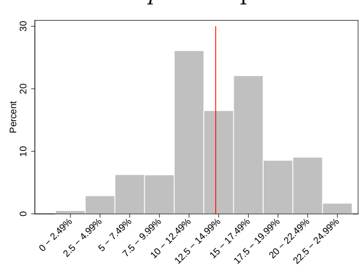
a.4. p: Group 4



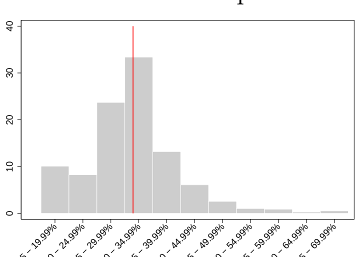
b.4. θ: Group 4



a.5. p: Group 5



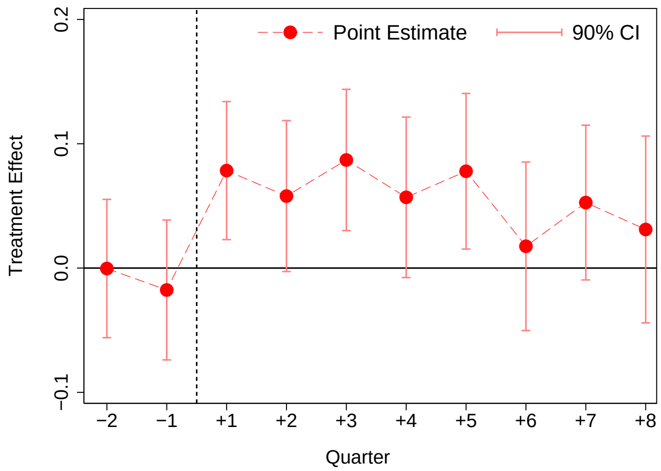
b.5. θ: Group 5



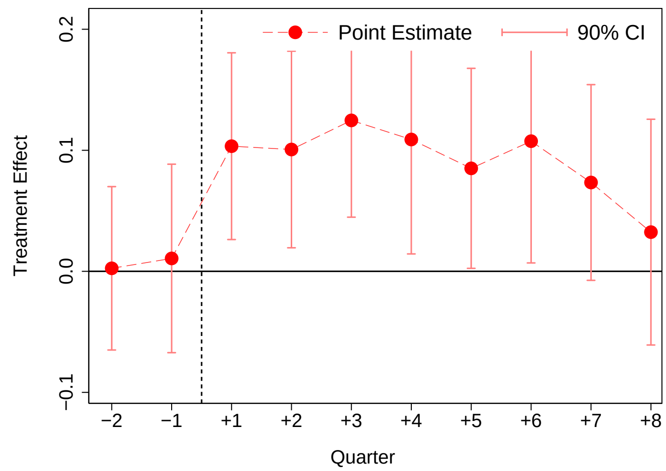
Notes: Notes: N=10,272. Information provided in the audit-statistics letter: probability of being audited (p, in panel a) and penalty rate (θ, in panel b). Group 1 through 5 correspond to each of the pre-treatment VAT payment quintiles.

Figure 2: Average Effects of *Audit-Statistics*, *Audit-Endogeneity* and *Public-Goods* Messages, By Quarter

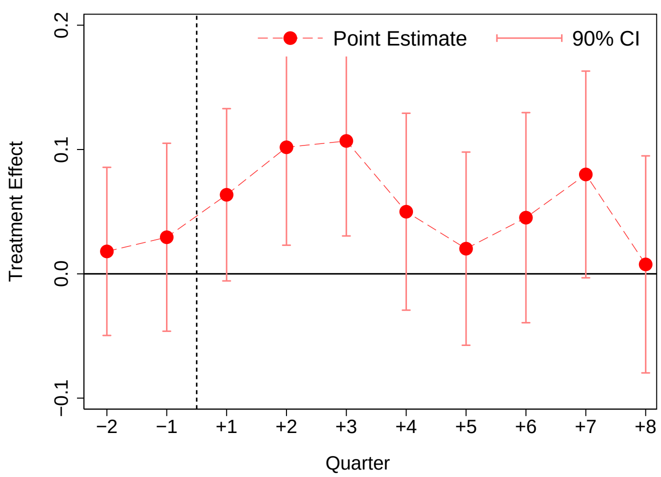
a. *Audit-Statistics* vs. *Baseline*



b. *Audit-Endogeneity* vs. *Baseline*

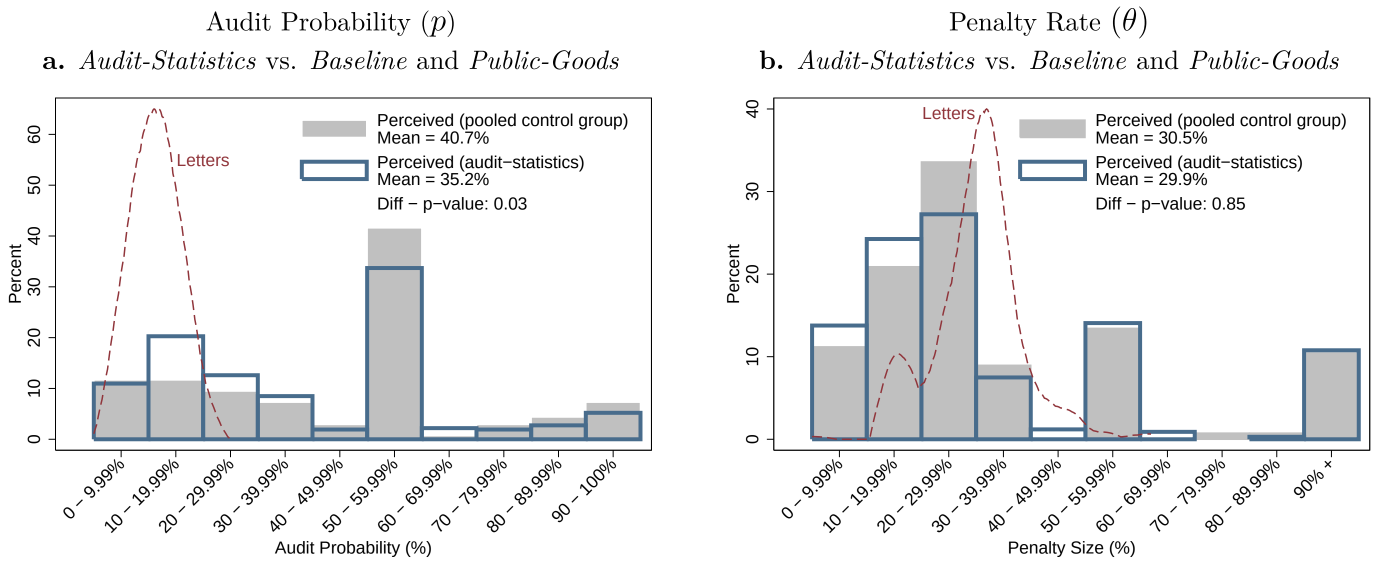


c. *Public-Goods* vs. *Baseline*



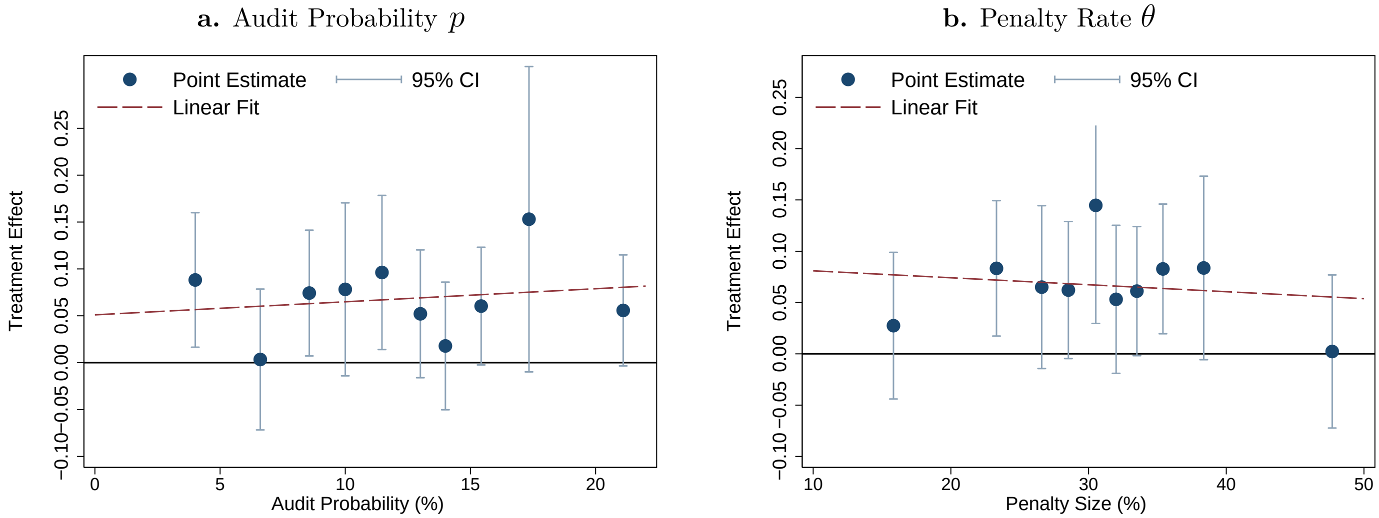
Notes: These figures plot the quarterly effects of each treatment arm when compared to the *baseline* letter. Panel a. (N=12,336) presents the effect of the *audit-statistics* message on total quarterly VAT payments, while panel b. (N=4,103) represents the effect of the *audit-endogeneity* message and panel c. (N=4,081) depicts the effect of the *public-goods* message on the same outcome variable. Each point (red circle) in the plot represents the estimate of the effect of treatment on VAT payments for a specific quarter from two quarters before treatment up to eight quarters after receiving the letter. Regressions do not include monthly pre-treatment controls. The results are based on Poisson regressions, so coefficients can be interpreted directly as semi-elasticities. 90% confidence intervals represented by red lines computed with robust standard errors.

Figure 3: Survey Results: Perception of Audit Probabilities and of Tax Evasion Penalty Rates by Treatment Group



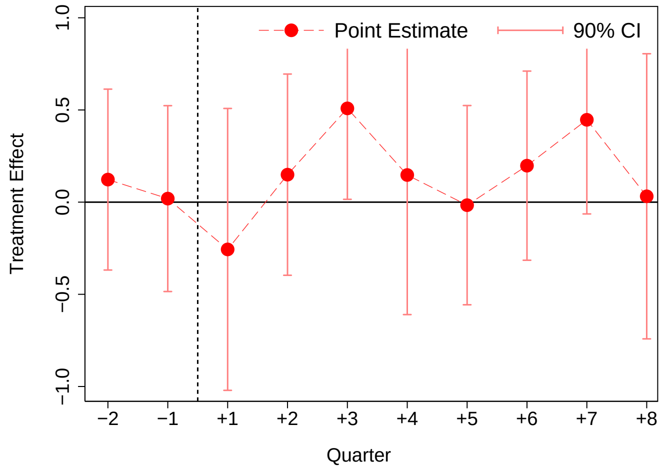
Notes: The histograms are based on survey responses from those who reported themselves as owners in the post-treatment survey. Perceived (pooled control group, N=137) refers to to survey respondents who received the *baseline* (N=69) or the *public-goods* (N=68) letters during the experimental stage (none of the two letters contained any information about audit probabilities nor penalty rates). Perceived (aud.–statistics) refers to respondents who received *audit-statistics* letters (N=365). In panel a. the x-axis represents the probability of being audited; in panel b. it represents the average penalty rate. We report the mean responses and the p-value of the difference between the two groups. The answers correspond to questions Q2 and Q4 (see full survey questionnaire in Appendix A.7). The red line represents the density function of the information displayed in the *audit-statistics* letters, measured in the right y-axis (hidden for the sake clarity).

Figure 4: Effect of *Audit-Statistics* vs. *Baseline* by Deciles of p and θ



Notes: Panel a. plots the first-year effect (October 2015 – September 2016) of the *audit-statistics* letter on total VAT payments by decile of p while panel b. reports the results from the same regressions by decile of θ (N=10,272). In both panels, each dot represents the estimated treatment effect for each decile of the parameter considered. Regressions are estimated using monthly pre-treatment controls and the stratification variable used to randomize the parameters. All effects are depicted with a 95% confidence interval. The results are based on Poisson regressions, so coefficients can be interpreted directly as semi-elasticities. Confidence intervals computed with robust standard errors.

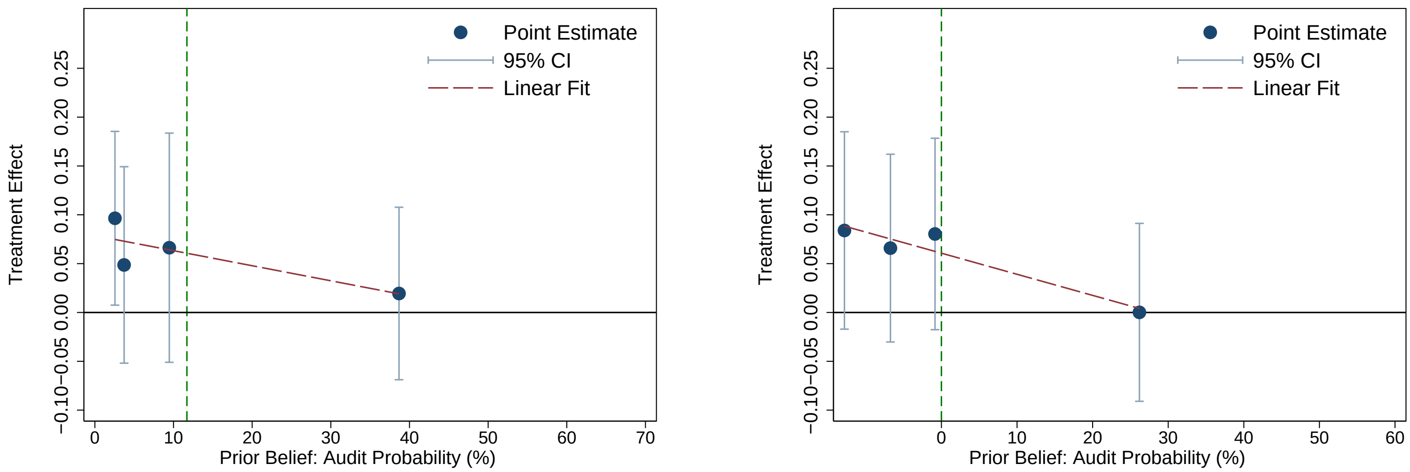
Figure 5: Effects of *Audit-Threat* Sub-Treatments: $p = 0.50$ vs. $p = 0.25$



Notes: This figure depicts the quarterly effect of the *audit-threat* message (N=4,048 and 50% vs 25%). Each dot (red circle) represents the estimate of the effect of treatment on VAT payments for a specific quarter from two quarters before treatment up to eight quarters after receiving the letter. Regressions do not include monthly pre-treatment controls. The results are based on Poisson regressions, so coefficients can be interpreted directly as semi-elasticities. 90% confidence intervals represented by red lines computed with robust standard errors.

Figure 6: Effect of *Audit-Statistics* vs. *Baseline* by Prior Beliefs

a. Audit probability - *prior (calibrated to 11.7%)* **b.** Audit probability - *prior - signal (calibrated to 11.7%)*



Notes: Panel a. plots the first-year effect (October 2015 – September 2016) of the *audit-statistics* letter on total VAT payments by quartiles of prior beliefs while panel b. reports the same results by quartiles of the difference between the prior belief and the signal sent in the *audit-statistics* message (N=11,989). The prior belief (before the experiment) is computed as $\hat{p}_i = 1 - \left(1 - \frac{\alpha_0 + N_i}{\alpha_0 + \beta_0 + T_i}\right)^3$
such that the mean prior belief about the probability of being audited at least once in the following three years matches the actual average probability observed in our sample. In panel b. the signal for the placebo group was randomly assigned using the same strategy that for the *audit-statistics* group. The red dashed line represents the linear fit corresponding to the four estimates . In panel a., the dashed green line represents the average perceived probability (11.7%). In panel b., the dashed green line represents the point in which prior belief and signal are equal. In both panels, each dot represents the estimated treatment effect for each quartile of the variable considered. Regressions are estimated using monthly pre-treatment controls. All effects are depicted with 95% confidence interval. The results are based on Poisson regressions, so coefficients can be interpreted directly as semi-elasticities. Confidence intervals are computed with robust standard errors.

Figure 7: Mass Advertising Campaign Sample: Billboard Poster, United Kingdom's Tax Authority, 2012

Notes: Advertising campaign by the United Kingdom's tax authority, HMRC (Her Majesty's Revenue and Customs), 2012. Previously hosted at http://www.gov.uk/sortmytax (no longer available, accessed through http://web.archive.org/)