

How to Improve Tax Compliance? Evidence from Population-
wide Experiments in Belgium

Jan-Emmanuel De Neve, Clement Imbert, Johannes Spinnewijn,
Teodora Tsankova & Maarten Luts

May 2019

No: 1194

This paper has been updated & published in [the Journal of Political Economy](#)

Warwick Economics Research Papers

ISSN 2059-4283 (online)

ISSN 0083-7350 (print)

How to Improve Tax Compliance? Evidence from Population-wide Experiments in Belgium

Jan-Emmanuel De Neve

University of Oxford

Clement Imbert

University of Warwick

Johannes Spinnewijn

London School of Economics

Teodora Tsankova

University of Warwick

Maarten Luts*

FPS Finance

May 5, 2019

Abstract

We study the impact of deterrence, tax morale, and simplifying information on tax compliance. We ran five experiments spanning the tax process which varied the communication of the tax administration with all income taxpayers in Belgium. A consistent picture emerges across experiments: (i) simplifying communication increases compliance, (ii) deterrence messages have an additional positive effect, (iii) invoking tax morale is not effective. Even tax morale messages that improve knowledge and appreciation of public services do not raise compliance. In fact, heterogeneity analysis with causal forests shows that tax morale treatments backfire for most taxpayers. In contrast, simplification has large positive effects on compliance, which diminish over time due to follow-up enforcement. A discontinuity in enforcement intensity, combined with the experimental variation, allows us to compare simplification with standard enforcement measures. Simplification is far more cost-effective, allowing for substantial savings on enforcement costs, and also improves compliance in the next tax cycle.

Keywords: Tax Compliance, Field Experiments, Simplification, Enforcement

JEL-codes: C93, D91, H20

*We thank Johannes Abeler, Pedro Bordalo, Anne Brockmeyer, Stefano Caria, Clément De Chaisemartin, Michael Devereux, Peter John, John List, Robert Metcalfe, Karl Overdick, Ricardo Perez-Truglia, Emmanuel Saez, Joel Slemrod, Dmitry Taubinsky and seminar participants at Chicago, Oxford, Warwick, Antwerp, HMRC, National Tax Association, IMEBESS, IIPF, and The Behavioural Insights Team for helpful comments and discussions. We are grateful for the opportunity to design and run field experiments in collaboration with Belgium's Federal Public Service (FPS) Finance. Pre-analysis plans were recorded on the AEA RCT Registry with IDs AEARCTR-0000827 and AEARCTR-0002196.

1 Introduction

Tax compliance sits at the heart of the healthy functioning of societies. It is therefore of little surprise that gaining a robust understanding of the drivers of tax compliance is an important topic in the economics literature. Tax compliance involves both the truthful reporting of taxable income and the timely payment of tax dues. The growth in third-party reporting of income has limited the ability to misreport income (see [Kleven et al. \(2011, 2016\)](#); [Jensen \(2019\)](#)).¹ Tax administrations, however, continue to devote considerable resources to the collection of taxes. In the United States the annual cost of non-compliance with individual income taxes due to nonfiling, underreporting, and underpayment is estimated to total about \$319 billion ([Internal Revenue Service, 2016](#)). Closing the “tax gap” is a key objective for governments around the world and relies on improving our understanding of the drivers of tax compliance and the cost effectiveness of further interventions ([OECD, 2010](#); [HM Revenue & Customs, 2018](#)).

The classic work by [Allingham and Sandmo \(1972\)](#) provided a work-horse model for understanding tax compliance through pecuniary incentives that deter non-compliance. Since then, a large body of research has stressed the role of non-pecuniary motives more broadly (e.g., [Kirchler \(2007\)](#); [Luttmer and Singhal \(2014\)](#); [Besley et al. \(2019\)](#)), often referred to as tax morale. Recent work has also highlighted the importance of information frictions and private costs underlying tax compliance (e.g., [Bhargava and Manoli \(2015\)](#); [Hoopes et al. \(2015\)](#); [Benzarti \(2017\)](#)). There is now scattered evidence for these different drivers of tax compliance to be important across a variety of settings (see [Slemrod \(2018\)](#)), but several questions remain unanswered. How important are these different drivers relative to each other in the same context? Do their effects depend on the stage of the tax cycle? On the treated population? Do they interact or persist over time? Can they be leveraged to complement standard enforcement measures?

This paper aims to provide a comprehensive comparison of these three key drivers of tax compliance – deterrence, tax morale and information frictions – and also compares their effect to standard enforcement measures. We study compliance effects throughout the tax process – including the timing of tax filing, the reporting of taxable income, and the payment of taxes – for all individuals subject to personal income taxation in Belgium. We compare these potential drivers of tax compliance in the same context and put them at equal footing by varying the content of the tax letters sent by the Belgian tax authority (Federal Public Service Finance, FPS Finance). In total, we ran five population-wide experiments in

¹Recent empirical work investigates the misreporting of foreign income in developing countries (e.g., [Alstadsæter et al. \(2018\)](#)) and of taxable income in developing countries (e.g., [Pomeranz \(2015\)](#); [Naritomi \(2018\)](#)) where paper trails are missing or the enforcement capacity falls short.

collaboration with the FPS Finance over the course of three fiscal years, 2014-2016. This comprehensive approach allows us to replicate findings at different stages of the tax process and across fiscal years, and to estimate longer-term, repetition and interaction effects.

The standard communication from the tax administration to taxpayers consists of the request to file a tax return and the request to pay taxes. Follow-up correspondence takes place in the event of taxpayers being either late in filing their tax return or being late in paying their tax dues. Our main focus is to leverage these different phases of communication in order to simultaneously test varying treatments related to the simplification of information, the use of deterrence, or the appeal to tax morale. The simplification treatments entail shortening the length of the letters, reducing the information overload and highlighting the action-relevant information to the taxpayer. The deterrence treatments add a message to the simplified letter that makes the financial penalties explicit and/or highlight the enforcement actions in case of (further) non-compliance. The tax morale treatments add a message that highlights the public good value of tax expenditures and/or the social norms attached to filing and paying taxes on time.

Our experiments provide very precise and consistent results across the tax process and the respective samples of taxpayers addressed. Simpler communication and deterrence messages significantly increase compliance, inducing people to file and pay their taxes sooner. The effects are substantial. Despite tax withholding one out of three tax payers has a positive outstanding balance on their tax bill, adding up to a total of 3.8 billion euros in 2016 (about 10 percent of personal income taxes). The simplification and deterrence treatment together increased timely payment after receiving a positive tax bill by 1.4% (resp. .51pp and .53pp). The compliance effects are particularly large for the reminder letters sent to the late tax payers and tax filers: the combined effect of simplification and deterrence on timely compliance is 26% (resp. 10.0pp and 1.2pp) for paying taxes and 17% (resp. 2.6pp and 2.8pp) for filing taxes. Overall, it is reducing information overload and making action-relevant information salient that seems particularly effective. Tax payers are also successfully induced to comply by making potential penalties and their enforcement explicit, and by the encouragement to pay immediately to avoid these penalties.

We study the full dynamics of the treatment effects on late payers, which diminish over time as the tax administration takes further enforcement measures (including imposing garnishments and sending bailiffs) to eventually reach close to full compliance. The treatment effects at the end of the tax cycle are 1.0pp, which is ten times smaller than the effect at the payment deadline. Still, the cost savings on follow-up enforcement imply a large return to the letter treatments. We exploit an enforcement discontinuity, combined with our experimental variation, to disentangle their respective effects. For the sample of late tax payers around

the enforcement cut-off, we estimate that the simplification treatment would have increased compliance by 9 percentage points in the absence of standard follow-up enforcement.

In contrast with the simplification and deterrence treatments, our treatments that seek to improve tax morale obtain no compliance effects and sometimes even reduce compliance. The ineffectiveness is replicated across all treatments arms, which include messages that invoke social norms, emphasize the social value of public expenditure or combine the two. We also experiment with a pie chart pop-up on how taxes are spent for online tax filers, which does not affect reported taxable income or claimed tax deductions, nor the perceived importance of honesty as measured in an endline survey. The survey suggests that while the tax morale treatment does not trigger a shift in compliance, it does increase taxpayers' knowledge and appreciation of public services.

Our setting allows us to push the frontier on the evaluation of letter treatments and behavioral interventions more generally in three important ways:

First, while nudges are by definition low-cost interventions, a key challenge is to know how they compare to the standard policy levers that they complement ([Benartzi et al., 2017](#)). To address this, we use the enforcement discontinuity to compare the causal impact of regular enforcement interventions to the experimental letter treatments for the exact same people (i.e., late taxpayers around the enforcement threshold). Projected on the sample of late taxpayers, a back-of-the envelope calculation tells us that the simplification treatment for this experiment alone could have increased tax collection by €20.2 million, or alternatively, amounted to savings on enforcement costs worth €5.4 million. The implementation costs of the nudge intervention were trivial in comparison (€79,511).

Another important concern is whether the gains from nudges are long-lived ([Allcott and Rogers, 2014](#); [Cronqvist et al., 2018](#)): do one-time interventions have long-term effects? Do repeated interventions remain as effective? We repeated the experiment on the late taxpayers in two consecutive years. We find that there are no diminishing marginal returns to repeating the treatment in that recidivists are equally responsive to a simplified letter independent of the letter type they received in the previous year. Moreover, we find that the effects extend to the following fiscal years: late payers are less likely to be late again in the next year after having received a simplified reminder letter in the first year, but this effect is offset if they received a tax morale treatment as well. In fact, the tax morale treatment has a significant, negative impact even two years later.²

Finally, the population-wide nature of our experiments give us sufficient power to study

²These findings extend on [Brockmeyer et al. \(2019\)](#), who find sustained effects from a deterrence message on firms' tax compliance in Costa Rica. These findings differ from [Guyton et al. \(2016\)](#), who find no long-term effects and positive returns from repeating reminders in claiming EITC.

heterogeneity in treatment effects, which are key to think about the distributional welfare consequences, but can also be leveraged to target the interventions where they are most effective. We use machine-learning techniques (Wager and Athey, 2018; Chernozhukov et al., 2018) to study heterogeneous treatment effects based on observables in the late payers experiment. While the dispersion is substantial, the simplification treatment effects are always positive. The deterrence treatments are unlikely to backfire, while the tax morale treatments almost never increase compliance. We also find that simplification was most effective among taxpayers with children, potentially consistent with their reduced attention span. The treatment effects are also smaller among tax payers with large outstanding tax liability, while we find no discernible differences by income. Finally, the tax administration predicts the solvency of all tax payers and we find that the treatments are the least effective among taxpayers with either very low or very high solvency scores.

Our paper contributes to the long literature studying the drivers of tax compliance and the growing number of randomized controlled tax trials in particular (see Slemrod (2018)).³ The key advantage of our experimental setting is that we test the main drivers of tax compliance in the same way, in the same setting, and on the same sample, which ensures comparability. Moreover, we intervene through the standard communication by the tax authority to its tax payers, but do so at different stages of the tax process and for different subsets of the tax payer population, which gives some external validity to our design. We also test different variations of similar treatments and study heterogeneous treatment effects, both of which help with establishing robustness and uncovering underlying mechanisms. The advantages of our setting are particularly valuable when results in the literature are mixed, as is the case for interventions appealing to tax morale. For example, Del Carpio (2014) finds positive and long-lasting impacts from invoking social norms on compliance in property taxation in Peru. Hallsworth et al. (2017) find that social norms and public services messages in official reminder letters increased payment rates for overdue tax in the UK. Bott et al. (2017) find that the inclusion of a moral appeal increases the reporting of foreign income in Norway. However, several other experiments testing normative appeals have found null or even negative results (e.g., Blumenthal et al. (2001), John and Blume (2018)). In particular, Cranor et al. (2018) test both deterrence and social norm treatments on the compliance by late tax payers in Colorado. They find that invoking social norms has no compliance effects, while making the penalty explicit has.⁴ Among the behavioral drivers of tax compliance there is an

³On the role of enforcement and deterrence, see reviews by Andreoni et al. (1998) and Slemrod and Yitzhaki (2002). An example of an RCT changing audit probabilities is Kleven et al. (2011). An example of an RCT changing the penalty information is Cranor et al. (2018). On the psychological, cultural, social, and normative factors underlying tax compliance, see reviews by Torgler (2007), Alm (2012), and Luttmer and Singhal (2014).

⁴Perez-Truglia and Troiano (2018) find that shaming tax payers by making their non-compliance public

increasing focus on the role of simplification. While this paper does not address tax system complexity by itself and how that affects taxable behavior (e.g., [Chetty and Saez \(2013\)](#), [Abeler and Jäger \(2015\)](#), [Aghion et al. \(2017\)](#)), it does shed new light on the informational complexity that is often associated with the process of filing and paying taxes and claiming benefits (see e.g., [Slemrod et al. \(2001\)](#), [Kleven and Kopczuk \(2011\)](#), [Hoopes et al. \(2015\)](#), [Cox et al. \(2018\)](#)). [Bhargava and Manoli \(2015\)](#) identify psychological barriers to the take-up of EITC benefits due to information complexity – with the mere simplification of the mailing leading to a significant increase in take-up. In a similar spirit, we show that simplifying the communication of the tax administration has a substantial effect on tax compliance – and that this effect can outweigh the effects of nudges related to deterrence and tax morale.

The paper proceeds as follows. We describe the context and empirical setting in the next section. Section 3 discusses the main experimental results by treatment categories, while Section 4 digs deeper within treatment categories to shed some light on mechanisms. Section 5 analyzes the regression-discontinuity in enforcement, compares its relative cost-effectiveness and discusses welfare and heterogeneous treatment effects. Section 6 concludes.

2 Context and Design

This section presents the five experiments we study and describes the experimental samples. We also provide some background on the tax filing and payment cycle for personal income taxation in Belgium.

2.1 Tax Process

The context of our study is Belgium, which in 2017 had a tax-to-GDP ratio of 44.6% (higher than the OECD average of 34.2%). We focus on individual income tax, the largest source of tax revenues. In the fiscal year 2016, individual income tax raised 27.7% of overall tax revenues, from 7.1 million taxpayers. Income taxes are collected solely at the federal level. There is a personal tax-free allowance which is currently at 12,990 EUR and marginal taxes rise from 25 to 50%.⁵ Fiscal years run from January 1st to December 31st, and the tax cycle starts in July of the year after the fiscal year in which the income has been earned. There are four main steps in the annual personal income tax cycle, as shown in Figure 1a: tax filing, filing reminders, tax payment and payment reminders. We vary the correspondence

increases compliance. However, they find no effects from providing information on others' non-compliance.

⁵In comparison, in the US, the tax-to-GDP ratio is lower (27.1%) and income taxes are more important as a share of tax revenues (38.6%). Federal marginal tax rates are lower (10 to 37%), but lower levels of government levy additional taxes.

between the tax administration and taxpayer at each of these steps.

Tax filing (TF): Taxpayers can file their taxes on paper or online, either by themselves or with the help of an accountant or a tax official.⁶ The online portal called “Tax-on-Web” is increasingly popular and in 2017 it was used by 3.8 million taxpayers, of which 1.7 million submitted their declarations individually, and the rest filed with the help of an accountant or a government official (we exclude them from the analysis).

Filing reminders (TFR): Figure 1b depicts what happens when taxpayers miss the filing deadline. Filers who have not submitted by the deadline are sent a filing reminder letter, and given 14 days to file. If a taxpayer has still not filed seven days after this second deadline, the tax administration uses its own estimates to compute their tax liability. In the fiscal year 2016, about 170,000 taxpayers had not filed by the deadline, which represents about 3.5% of taxpayers who were expected to file.

Tax payment (TP): A majority of taxpayers are taxed at the source if they are employees or pre-pay their taxes based on estimates of their tax liability if they are self-employed. A significant share of taxpayers also have taxable income below the exemption threshold and thus pay no income taxes. As a result, less than a third of taxpayers (1.9 million in the fiscal year 2016) receives a tax bill with a positive payable balance, which they need to pay within the next two months. The majority of such cases can be explained by insufficient withholding at the source in situations that made it difficult to calculate the exact tax liability (e.g. tax payers who hold several jobs, students who work part-time, etc.). Total taxes due at that stage are 3.8 billion euros.

Payment reminders (TPR): Figure 1c depicts what happens when taxpayers miss the payment deadline. Taxpayers who have not paid two months after receipt of the tax bill are sent a payment reminder. Taxpayers who still do not comply are then exposed to further enforcement actions, which start after 14 days. In the fiscal year 2016, about 220,000 taxpayers had still not paid 14 days after the deadline, and owed a total of 0.8 billion euros, which represents 12% of taxpayers who received a positive tax bill, and 21% of taxes they owed.

⁶Not all taxpayers need to file. About a third of taxpayers (2.2 million in the fiscal year 2016) receive pre-filled tax returns with no further action required.

2.2 Experiments

We report on a total of five experiments: one on tax filing (TF), one on filing reminders (TFR), one on tax payment (TP) and two on payment reminders (TPR). The experiments spanned three fiscal years (FY), FY2014 to FY2016. The experiments randomized different treatments that we categorize in three groups: simplification, deterrence and tax morale.

In four experiments out of five, the treatment involved simplifying the letter to communicate what the tax administration expected from taxpayers. Simplification included shortening the letter while retaining the action-relevant information. To attract the attention of the reader, important information was highlighted in color and/or placed in boxes. The exact implementation of the treatment varies across treatments as we discuss below. In most cases, the letter was also personalized, i.e. it was addressed to the taxpayer using his/her name. The French versions of the letters are shown in Appendix A.7 to A.12; Flemish and German versions were sent to Flemish and German speaking taxpayers, respectively.

The experiments also tested the effect of deterrence and tax morale through the addition of short messages in the simplified letter. The deterrence messages aimed at making the consequences of non-compliance explicit, by explicitly stating fines and tax increases and/or by explicitly mentioning follow-up enforcement. We also tested messages that encouraged immediate action to avoid the fines. The tax morale messages, on the other hand, aimed at raising compliance by increasing the desire of taxpayers to comply with social norms or to reciprocate for public goods provision. Appendix Table A.1 lists the deterrence and tax messages used (translated in English).

TP Experiment: The Tax Payment experiment modified the tax bill sent to taxpayers with a positive liability: the experiment was carried out between November 2017 and May 2018 with 1,216,317 taxpayers (fiscal year 2016). All treated taxpayers received a simplified letter, only keeping action-relevant information and improving the overall outline: Appendix Figure A.7 shows the old letter, and Appendix Figure A.8 the simplified letter. For a subset of treated individuals, the letter included either deterrence messages or tax morale messages (see Panel A of Appendix Table A.1). For this experiment, outcomes include the probability of making a payment following letter receipt (extensive margin response), and the fraction paid conditional on a payment having been made (intensive margin). As baseline outcome, we use the probability of payment within 60 days after the letter was sent: 60 days is the deadline given to taxpayers to pay their outstanding debt.

TPR Experiments: The Payment Reminder experiments were conducted with taxpayers who were late in paying their tax: 229,751 taxpayers in 2015/16 (FY2014) and 188,180

taxpayers in 2016/17 (FY2015).⁷ The treatment group received a simplified reminder letter, in which the outstanding tax liability and the deadline were highlighted and other information shortened: Appendix Figure A.9 shows the old letter, and Appendix Figure A.10 the simplified letter. Again, for different subsets of the treatment group, the letter also included deterrence and tax morale messages (see Panel B of Appendix Table A.1). The baseline outcome we consider is now the probability of payment within 14 and 180 days after reminder receipt: 14 days corresponds to the time at which enforcement actions begin. To validate results and to test the effect of repeated treatments, the TPR experiment was conducted in two consecutive years.

TF Experiment: The Tax Filing experiment was conducted in 2017 (FY2016) with 1.5 million online tax filers.⁸ The tax filers were shown a pop-up pie chart either before (treatment) or after (control group) they filed their taxes. The pie chart presented the breakdown of government spending by categories (see Appendix Figure A.1).⁹ The chart was accompanied by a sentence highlighting that these public services were funded by taxes.¹⁰ We consider this as a similar treatment to the tax morale message in the other experiments. For this experiment, outcomes come from two sources: administrative data on tax compliance and answers to an online survey to which all online filers were invited. Due to confidentiality concerns, the administration did not provide individual information but only average outcomes (or taxpayers characteristics) within a gender-age cell. The main compliance outcome is reported taxable income. Other outcomes are tax liability, self-employed profits and expenses, expenses of salaried workers and general expenses. These are also based on declared values. Survey data is available for those who agreed to answer the questionnaire, which gauges taxpayers' knowledge and agreement with the way tax revenue is spent, and their evaluation of public services and the tax system more generally. The survey instrument is described in Appendix A.2.¹¹

⁷In both trials, German speaking taxpayers, taxpayers who had raised objections to the amount they owe in unpaid taxes and individuals with a BIS number for whom the government did not have a name were not included in the randomization and received an old letter. Only debts related to the current fiscal year and letters that are first means of communication with the taxpayer (no updates on balances owed) are included in the analysis.

⁸This excludes taxpayers who used an accountant or tax officer to submit their taxes via the online portal. Our dataset covers taxpayers who submitted their tax returns before mid-August 2017.

⁹The tax administration also provided a pie chart of government expenditures by region, which was available when scrolling down.

¹⁰For some randomly selected sub-groups, the administration added at the very bottom of the pop-up an additional sentence that either added a public goods message, mentioned penalties in general terms, or appealed to social norms in general terms (see Panel C of Appendix Table A.1). We do not find any differential effect of this second sentence and pool all treatment groups in the analysis.

¹¹All outcome variables were pre-specified in the Pre-analysis Plan (AEARCTR-0002196).

TFR Experiment: The Filing Reminders experiment was conducted with 148,925 taxpayers who were late in filing their tax returns in 2016 (FY2015). The treatment group received a simplified letter, which emphasized the new filing deadline: Appendix Figure A.11 shows the old two-page long letter and Appendix Figure A.12 shows the one-page simplified letter. A subset of the treatment group received a letter which included deterrence messages (see Panel D of Appendix Table A.1).¹² For these experiments, the baseline outcome is the probability of filing within 21 days after letter receipt: 21 days is the time at which the tax administration begins to calculate the tax liability based on income estimates.

2.3 Randomization Design

The allocation of taxpayers to the different treatment groups was done in two different ways. For the TPR, the TF and the TFR experiments, it was based on the last two digits of the national identity number, which are random. For the TP experiment, treatment allocation was based on the day of the month the taxpayer was born, which is also random and independent of the last digits of the national identity number. Appendix Table A.2 displays the assignment of the two last digits of the national identity number to the different treatment groups for the TFR, TPR, and TF experiments. There are three things to note.

First, treatment allocation for the TPR 2014 experiment was done on the basis of the penultimate digit, while the allocation for the TPR 2015 experiment was done on the basis of the last digit only (see Appendix Table A.3). This implies that the two allocations are independent from each other, as in a cross-cutting randomization design. As there is significant overlap between 2014 and 2015 late payers (see Appendix Table A.4), we can estimate the effect of the two treatments both separately and jointly, to identify the effect of repeated treatment.

Second, treatment allocations for the TPR 2014 and TFR 2015 experiments coincide partially, but not completely. A potential concern could be that treatment status in one experiment affects outcomes in a following experiment. Fortunately, the two experiments were done on different target populations, since the late payers of 2014 need not be late filers in 2015. Indeed, the overlap between the two populations is small: as Appendix Table A.4 shows, only 6% of late payers for the fiscal year 2014 were also late filers for the fiscal year 2015. As a robustness check, we estimate the results of the TFR 2015 experiment controlling for the TPR 2014 treatment assignment and show that our results do not change.

Third, treatment allocation for the TF 2016 experiment again split the tax sample in two

¹²In the previous year (FY2014), the administration carried out a separate experiment on filing reminders, in which it included tax morale messages without simplifying the letter first. We managed to collect data from this experiment and found no effect of the treatment. See Appendix A.1 for more details.

based on the two last digits of the national identity number, which made it partly, but not completely coincide with treatment allocations for the TFR and the TPR 2014 experiments. Unfortunately, to protect privacy the tax administration did not share individual identifiers for the TF 2016 experiment, so we cannot measure the exact overlap with the sample of the other two experiments, or control for assignment to previous treatments. However, since the sample of the TF experiment is much larger (1.5 million, against 150,000 for TFR and 230,000 for TPR 2014), the overlap is likely to be small.

2.4 Population comparison

By focusing on different stages of the tax process, the five experiments test the effect of all treatment categories on different parts of the taxpayer population. Table 1 shows descriptive statistics on socio-demographic characteristics of the different experimental samples, as compared to the universe of Belgian taxpayers.

The Belgian personal income taxpayer is on average 49 years old, in a couple in 35% of the cases and has 0.4 children (column 1). By convention, in the case of households composed of individuals of both genders, only the gender of the woman is recorded, so that there are many more female than males (70%). 33% of the taxpayer population lives in Wallonia and 42% speak French. On average, they owe €570, but only 28% have a positive tax liability. Taxpayers in the TP experiment have a tax liability which is by definition positive, with an average of €2676. As column 2 shows, they are older, more likely to be in a couple and less likely to have children. In contrast, taxpayers in the TF experiment (column 4) are the online filers: they are younger, and have more children. Taxpayers in the reminder experiments (TPR and TFR in columns 3 and 5) differ from the overall population in similar ways: they are more likely to be male, less likely to be in a couple, younger, more likely to speak French and to live in Wallonia. In addition, the taxpayers who are late in paying (column 3) have lower tax liability than the average taxpayer with positive liability (€1890).

3 Experimental Results

3.1 Baseline Results

To estimate the effect of complexity, deterrence and tax morale throughout the four experiments, we exploit the randomization and simply regress compliance outcomes on treatment dummies and taxpayer controls. The estimating equation writes:

$$Y_i = \alpha + \beta_S S_i + \sum_j \beta_j T_i^j + \gamma X_i + \varepsilon_i,$$

where Y_i is the relevant outcome for taxpayer i , S_i is a dummy variable equal to one for taxpayers who receive a simplified letter, T_i^j are dummy variables equal to one for the different types of messages, and X_i denotes a vector of taxpayer controls.

The outcome variable Y_i we use for our baseline specification in the tax payment experiment is whether the tax liability (in full or in part) is paid before the deadline, which is 60 days after the letter receipt. For the reminder experiments, the outcome variable is whether taxes are filed or paid before the start of follow-up interventions (respectively after 21 and 14 days for the filing and payment experiments). We consider compliance at different horizons and at the extensive vs. intensive margin below. For the tax filing experiment, the compliance variable is different in nature, as we look at reported taxable income. Controls X_i include dummies for age, region, mother tongue, number of children as well as dummies for time at which the letter was sent for experiments in which letters were sent out in waves. There are some differences in the remaining controls across experiments; the full list can be found in Table 1¹³.

The coefficients of interest are β_S , which identifies the effect of simplification, and β_j , which identifies the additional effect of deterrence or tax morale. As described in the previous section, there was no tax morale treatment in the filing reminder experiment.

Figure 2 presents our baseline estimates for the simplification, deterrence and tax morale treatment. The tax payment and tax filing experiments are in the top and bottom panels respectively. The experiments on the baseline sample of tax payers/filers are on the left, while reminder experiments for the late payers/filers are on the right. The figure conveys a very clear and strong pattern across the four experiments. In the three experiments in which communication with the taxpayer was simplified (TP, TPR and TFR), it had a positive effect on tax compliance. In the same three experiments, the deterrence messages had an additional positive, significant effect. Finally, in the three experiments in which the administration tried to increase tax morale (TP, TPR and TF), it had either no effect or even reduced compliance.

The regression estimates are also presented in Table 2, which has the same structure as Figure 2. The top panel (Panel A) presents the results of the tax payment experiments. Column 1 shows that as compared to the control group, in which 72.8% of taxpayers paid their taxes on time, simplifying the tax bill had a positive effect on the probability of paying on time, increasing it by 0.51pp (se .15pp). Adding a deterrence message increased the prob-

¹³In the payment reminders experiments we include also dummies for amount owed, income and solvency score quintiles, in the tax payment experiment dummies for amount owed quintiles and in the tax filing experiment dummies for marital status categories. The survey data is not linkeable to the admin data for confidentiality reasons, but includes information on the last two digits of the tax ID (used for randomization), gender and age.

ability of paying on time further, by 0.53pp (se .11pp). These effects are relatively small, but significant. The tax morale messages, however, had no additional effect on tax compliance. The effect of -0.12 pp (se .09pp) is sufficiently precisely estimated to rule out effects of a magnitude comparable to the simplification and deterrence treatment. Column 2 presents the results of the payment reminders experiment. The results are qualitatively similar and the effects of simplification and deterrence are again substantial. That is, simplifying the reminder letters increased the probability of paying by 10pp (23% of the control mean), and deterrence messages had an additional positive effect of 1.2pp (3% of the control mean). Tax morale messages, however, had an opposite effect, slightly reducing tax compliance (-0.7 pp). The bottom panel (Panel B) presents the results of the tax filing experiments, which are again very similar qualitatively. The tax morale treatment in the tax filing experiment has no effect on declared taxable income (an effect of .0011 on log income), with the null effect again being precisely estimated (se .0013). The estimates in Column 2 of Panel B show that simplification and deterrence had a large positive effect on tax compliance among late filers. Those who received a simplified letter were 2.6pp more likely to file on time. This probability increased by an additional 2.8pp for those who received a simplified letter with a deterrence message, making them 17% more likely to file on time compared to the control group.¹⁴

3.2 Dynamic Effects

We have reported treatment effects at one point in time, respectively at the deadline (TP) and before the start of enforcement actions (TPR, TFR). Using the payment and filing history, we can estimate treatment effects at any time – measured in days – after treatment. Let $Y_{i,t}$ be the tax compliance outcome of individual i at time t . As before, S_i denotes a dummy variable equal to one for taxpayers who received a simplified letter, and T_i^j are treatment dummies for deterrence and tax morale messages. We estimate the following equation:

$$Y_{i,t} = \alpha_t + \beta_{S,t}S_i + \sum_j \beta_{j,t}T_i^j + \gamma X_i + \epsilon_i.$$

For the tax payment experiment, t ranges from the receipt of the tax bill to 60 days after, corresponding to the deadline. For the payment reminder experiment, t ranges from the receipt of the letter to 180 days after. Note that the deadline was two days after, and that enforcement follow-up does not start until fourteen days after. For the filing reminder

¹⁴Appendix Table A.13 presents the results of the filing reminder experiment controlling for the treatment assignment in the payment reminder experiment. Due to the partial overlap between the two experiments, the estimates are less precise, but the treatment effects are very similar.

experiment, t ranges from the receipt of the letter to 60 days after, when for the majority of non-compliers automatic filing by the administration had finished. The deadline for the late filers is at 14 days.

Appendix Figure A.2 displays the dynamics of tax compliance in the control group - the estimated α_t - for the three experiments. In the TP experiment, the proportion of taxpayers who paid in the control group increased slowly after receipt of the tax bill, and then sharply just before the deadline, so that 72% of taxpayers met the deadline. In the TPR experiment, only a minority of late payers (17%) met the renewed deadline, and less than half of them had paid before the beginning of enforcement actions. The pattern is similar in the TFR experiment: only 25% of late filers in the control group had filed by the renewed deadline and only 34% had filed before enforcement actions began.

Figure 3 presents the dynamics of the simplification treatment, $\beta_{S,t}$. Taxpayers who received a simplified tax bill were slightly more likely to pay in the first weeks after tax bill receipt, but the difference with the control group really widened in the last week before the deadline. This suggests that simplification made both the need to pay and the actual deadline more salient to taxpayers. For the late payers, the simplified reminders had a strong and immediate effect on payment probability, which peaked around the time when enforcement actions started. As enforcement actions began, the control group caught up with treatment, so that the treatment effects decreased steeply, from above 10pp to below 1pp after six months, although they were still statistically significant at the end of the period. In Section 5, we disentangle the compliance effect of the simplification treatment and the follow-up interventions. In the filing reminder experiment, the simplified reminders also had a strong and rapid effect on filing probability, which again peaked at the time at which enforcement actions started. Then, as income was automatically filed, the difference in manual filing remained constant between treatment and control. For completeness, we report the dynamics of the effects of deterrence and tax morale messages, $\beta_{j,t}$, in Appendix Figure A.3. Across the three experiments, the additional positive effect of deterrence messages, which emphasized the penalties associated with missing the deadline, were felt gradually, and peaked at the deadline. In the Payment Reminder experiment, the negative effect of tax morale messages lingered for about a month, even after enforcement actions begun.

3.3 Long-term Impact and Repeated Treatment

We now turn to the question of whether the impact of the nudge intervention was short-lived. We first investigate whether one-time interventions can have long-term effects. We use the randomization in the FY2014 payment reminder experiment to estimate the effect

of reminder letters on timely payment in the next fiscal year. The results are shown in column 1 of Panel A of Table 3. We find a positive and significant effect of simplification on tax compliance in the next financial year. The probability of paying taxes on time in FY2015 increases by 1.1pp (se 0.28pp). Note that this long-term effect of simplification of the reminder letter is even slightly larger than short-run effect of the simplification of the tax bill itself. Of course, the reminder letters were sent to a subsample of taxpayers only. Moreover, as discussed below, the simplification of the reminder letter was more substantial than the simplification of the tax bill itself. In contrast with the simplification effects, deterrence messages have no additional positive impact in the following financial year, while tax morale messages have, if anything, a negative effect, almost entirely offsetting the long-term impact of simplification. Moreover, while simplification effects seem to get smaller over time, the tax morale treatment is associated with significantly lower compliance, even two years after letter receipt (column 2 Panel A of Table 3). Overall, these estimates suggest that small nudges can have long term effects.

We also test whether repeated interventions remain effective. We again study this for the payment reminder experiment, which was carried out for two consecutive years. That is, the experiment was run in FY2015 with the same treatments as in the main FY2014 experiment. Moreover, the treatment allocation followed two independent random numbers in the respective years. The repeated treatment allows us to check that our main results are replicated. The results of the FY2015 Payment Reminder experiment are shown in Panel B of Table 3. In FY2015 as in FY2014, simplifying tax reminders had a large positive effect on the probability of paying before enforcement starts (+25%), and deterrence messages had an additional positive effect (+3.8%), while tax morale messages had a negative effect (-2.7%). Interestingly, mixing deterrence and tax morale messages had a smaller impact than deterrence messages alone with a significant difference at 14 days.¹⁵

The cross-randomization of the two experiments further allows us to test whether simplified letters had a larger or smaller effect for taxpayers who received them twice, i.e. whether repetition induced a reinforcement or a fatigue effect. Only 28% of FY2014 late payers were late again in FY2015, so we lack power if we estimate FY2015 treatment effects on all FY2014 late payers. Instead, we use only taxpayers who were late twice, which is a selected sample, given the long-term effect of the FY2014 treatment. The selection effect can work against us in finding a positive treatment effect in FY2015, since taxpayers who are still late

¹⁵Note also that other nudges do not work in combination with non-simplified letter. In the case of Filing Reminders, an experiment was run in 2014, but unlike the main 2015 experiment, only tax morale messages were used, and without simplifying the letter. As Appendix Table A.8 shows, these messages had a null or negative effect on the probability of filing before enforcement actions started. These results confirm that tax morale messages do not improve tax compliance.

in FY2015 despite being treated in FY2014 likely have a lower propensity to pay. On the other hand, given that they are late again, the FY2014 treatment effect may have been particularly short-lived for these recidivists, making repetition of the treatment more effective. With this caveat in mind, we estimate the following equation:

$$Y_i^{2015} = \alpha + \beta^{2014} S_i^{2014} + \beta^{2015} S_i^{2015} + \delta S_i^{2014} \times S_i^{2015} + \gamma X_i + \varepsilon_i$$

where Y_i^{2015} denote compliance in FY2015, and S_i^{2014} and S_i^{2015} denote treatment assignment in FY2014 and FY2015 respectively. The results are presented in Panel C of Table 3. Among the selected sample of taxpayers who were late twice, simplification had a positive effect on payment in FY2015 (9.7pp), and was no less (and no more) effective for taxpayers who had already received a simplified letter in FY2014. The interaction effect δ is zero and relatively precisely estimated (se 1.2pp). We interpret our results as providing suggestive evidence against any fatigue effect, at least for recidivist tax payers.¹⁶

4 Robustness and Mechanisms

Our results are remarkably consistent across the four experiments, which were implemented at different stages of the tax process, and on different populations: simplification and deterrence have positive effects on tax compliance, while tax morale messages do not. This section discusses some further results by treatment category, aiming to convey the robustness of the results and hinting at some potential mechanisms. We focus on treatment variations within each category and their impact on tax compliance and other outcome variables.

Simplification Our experiments show that simplifying the tax correspondence can have a substantial impact on compliance. With a single experimental treatment, it is often challenging to establish the external validity of the results and to know which components of the treatment worked. The richness of our setting allows us to assess the robustness of the effect of the simplification treatment across experiments and to analyze the effect of slight variations in the treatment within an experiment.

A first step is to compare the magnitude of the effects of simplification across experiments. For this, it is important to keep in mind that while the simplified letters look very similar, the quality of the old letters was different. In particular, in the letter sending the tax bill, the required actions were already presented together and highlighted in the *old* letter, but made even more salient in the *new* letter (Appendix Figures A.7 and A.8). For the *old*

¹⁶Results for the deterrence and tax morale messages confirm this conclusion (see Table A.5).

payment reminder letter, the action-relevant information was hidden and spread out over a long, technical letter, also containing information that was only relevant for internal use (Appendix Figure A.9). The quality of the *old* filing reminder letter was arguably in between (Appendix Figure A.11). In the payment reminder experiment, the simplified presentation increased tax compliance by as much as 23% before the start of follow-up enforcement. This effect is larger than in the filing reminder experiment and an order of magnitude larger than in the payment experiment (see Table 2). Hence, simplification was effective everywhere, but had a larger impact in contexts where communication used to be more complex.

Next, for the tax payment experiments (TP and TPR), we can estimate the effects on both the extensive and intensive margins of tax payment. Panel A of Appendix Table A.7 shows the treatment effects on the fraction of the tax liability paid conditional on paying. We find positive effects of simplification at the intensive margin, but of much smaller magnitude than the extensive margin effects (and only significant in TP and TPR2015). Together with the dynamic patterns discussed before, with larger effects at the deadline and potentially right after receipt of the letter, the results suggest that the simplified communication is effective in making the deadline more salient and reduces chances to forget to pay or file before the deadline, but may also help overcome erroneous non-compliance and trigger more immediate action. Finally, the tax payment experiments (TP and TPR) also included treatments which varied more subtle features of the letter design, for example dropping the use of names in addressing the taxpayer and changing the order of the male and female partner of a couple. These treatments did not deliver any significantly different effect (see Table A.6).

Deterrence While prior work - both theoretical and empirical - has highlighted the importance of deterrence to tackle tax evasion, our experiments show that making penalties explicit in tax correspondence can increase tax compliance too. The effect is of similar magnitude across the different experiments, increasing compliance by around 1 – 2 percentage points. The baseline deterrence treatment in the tax payment and payment reminder experiments states the average penalty (of €209) explicitly. While personalizing stated penalty amounts is outside the scope of this study, the treatment effect is somewhat larger in the filing reminder treatment, in which instead of the average penalty, the range of possible penalties (from €5 to €1,250) and tax rate increases (from 10 to 200 percent) is stated. We also find that emphasizing the seizing of income/assets to collect penalties in addition to the stated penalty further increases compliance (from .1pp to 2.5pp in TPR, FY2015 - see Appendix Table A.6).¹⁷ We also tested a more implicit variation of this enforcement message,

¹⁷The difference in treatment effects between the explicit penalty and the explicit penalty+enforcement treatment is significant with a p-value of 0.001.

which is to emphasize instead that not paying taxes will be seen as an active choice, building on [Hallsworth et al. \(2015\)](#). This treatment had no significant effect, potentially in line with the ineffectiveness of the tax morale treatments in our context. In contrast, by emphasizing the returns to immediate action in terms of avoiding any potential penalty significantly increased compliance. In the payment reminder experiment, making the penalty explicit in combination with the message *“By paying now, you may still avoid these costs.”* increased compliance from 1pp to 1.9pp (see TPR, FY2015 in Appendix Table [A.6](#)).¹⁸ Also in the tax payment experiment, we have ran a treatment in which we explicitly highlighted the returns to immediate action, which increased the treatment effect from the simplified letter from .46 to 0.71 percentage points (see TP in Appendix Table [A.6](#)). This complements the earlier finding from the simplification treatment that besides making the relevant information salient, there is also a role for encouraging immediate action. We do not find an effect of deterrence at the intensive margin, when looking at the paid tax liability conditional on paying (Appendix Table [A.7](#)).

Tax Morale Our finding that tax morale messages are ineffective in raising tax compliance contrasts with some earlier studies on tax payment (e.g., [Hallsworth et al. \(2017\)](#) in the UK) and on tax filing (e.g., [Bott et al. \(2017\)](#) on foreign income reporting in Norway). However, a series of studies have found no effects when introducing normative appeals (e.g., [Blumenthal et al. \(2001\)](#), [John and Blume \(2018\)](#)). If anything, in our setting, invoking tax morale seems counterproductive: the negative effect is insignificant in the tax payment and tax filing experiment, but significant at the 5% and 1% level in the 2014 and 2015 payment reminder experiments, respectively. We both widen and strengthen the evidence by finding no or negative results at the payment and the filing stage, for the full population of tax payers / filers and on the subset of late filers / payers. Since we work on the universe of Belgium tax payers, the estimates are sufficiently precise to reject at usual significance levels that tax morale messages have effects of a magnitude comparable to the simplification and deterrence treatments. The tax morale message is also consistent across different treatment variations used in prior literature, either emphasizing the social value of the tax expenditures, or invoking the social norm of tax compliance by other Belgian taxpayers, framed in different ways, stated by themselves or interacted (see Appendix Table [A.8](#)). For the online tax filing experiment, the treatment is somewhat different (i.e., the pop-up of a pie chart of tax expenditures) and so is the compliance measure (i.e., reported taxable income). However, the conclusions are the same. Panel B of Appendix Table [A.7](#) shows the impact

¹⁸The difference in treatment effects between the explicit penalty and the explicit penalty+immediacy treatment is significant with a p-value of 0.047.

of the pie chart treatment on five other tax compliance outcomes, including self-employed profits and deductible expenses. The average treatment effect on tax compliance is precisely estimated, but always insignificant.

Survey Results To shed some light on the reasons why tax morale messages are ineffective, we can exploit the large-scale survey implemented in combination with the online tax filing experiment. Taxpayers who filed their taxes online for the fiscal year 2016 were invited to participate to the survey immediately after they filed. As part of the tax filing experiment, a random subset of them had seen the pie-chart describing government expenditures before they filed. The other tax filers had not yet seen the pie-chart, which was only shown to them after they had completed the survey (or declined to participate). A total of 79,334 tax filers completed the voluntary survey: take-up rates were similar in the treatment and control groups (resp. 5.15% and 5.14%).¹⁹ As an internal validity check, we first confirm that tax filers who were shown the pie chart were significantly more likely to state that they knew how taxes are spent. Respondents were also asked to state the actual expenditure share for each category. Using their responses we construct an index that is based on the standardized sum of absolute deviations between the stated and the actual share over all spending categories, and we find that treated tax filers scored significantly higher in that index (see columns 1 and 2 of Panel B of Table 4). At the very least, these results confirm that the pie chart was noticed and provided some new information. Second, we find that showing the pie chart significantly increased tax filers' agreement with how taxes were spent in general. Respondents were asked for their preferred ranking of categories of public spending in terms of budgetary priorities. Using survey responses and actual rankings, we construct a preference index and find that treated tax filers' stated preferences were closer to the actual expenditures (see columns 3 and 4). We also find that treated tax filers stated that they value the public services financed with tax revenues more. In the end, however, treated tax filers were not more likely to be satisfied with the general tax system and not more likely to agree with the statement that taxes should be reported honestly (see Panel A of Table 4). These results suggest that while the pie chart treatment was effective in improving taxpayers' knowledge and their appreciation of how their taxes were spent, this may have been insufficient to improve their tax morale.

¹⁹The only individual level characteristics available in the data sets are gender and age. We compare the treatment and control group within the pool of respondents and find a (small) difference in gender: respondents in the treatment group were 0.8% more likely to be male.

5 Enforcement, Welfare and Heterogeneity

This section compares the effectiveness of nudges in raising tax compliance among late tax payers with that of enforcement actions by the tax administration. We exploit a regression discontinuity in enforcement intensity, which, combined with the experimental design, provides a unique opportunity to estimate the causal impact of nudge-type interventions and of standard policy levers for the same population and in the same setting. We then discuss three different ways to assess the relative cost-effectiveness of nudges and standard enforcement actions. Finally, we explore heterogeneous treatment effects, which are indicative of the distributional effects of nudges and helpful to improve their targeting.

5.1 Nudges vs. Enforcement

The tax administration relies on various enforcement actions to make late payers comply. The first follow-up intervention for late tax filers and taxpayers is naturally the reminder letter, which we experimentally manipulated. Individuals who do not comply after receiving the reminder are subject to further enforcement actions. Local tax administrators have some discretion in the choice of enforcement mechanisms. Commonly used tools for payment non-compliers include sending registered letters (which require confirmation of receipt), imposing garnishments and the use of bailiffs. The dynamic pattern of the treatment effects (Figure 3) showed that the letter treatments accelerated tax payments, but that their final effect on tax compliance was more modest. The timing of the decline in treatment effects corresponds to the start of the enforcement actions undertaken by the administration, which suggests that these actions are responsible for the control group catching up with treatment.

To provide causal evidence on the effect of enforcement actions, we implement a regression discontinuity design which exploits exogenous variation in enforcement intensity at a specific threshold for the outstanding tax liability. We then combine the regression discontinuity with the simplification treatment to understand both how much the simplification treatment reduced the need for follow-up enforcement and how much the follow-up enforcement reduced the impact of the simplification treatment.

As Panel (a) of Figure 4 shows, there is a clear jump in the probability of enforcement actions above the tax liability threshold (normalized to 0 for confidentiality reasons), both in the treatment and control group.²⁰ There is no evidence of bunching below the threshold, which confirms that it is not known to the public (see Figure A.4). Moreover, before enforcement started, the probability of paying is smooth at the cut-off in both groups. This

²⁰We drop taxpayers with a liability exactly at the cut-off. The threshold value is a round number and the distribution of liabilities shows bunching at all round numbers in the vicinity of the threshold.

probability of paying, however, is much higher in the treatment than in the control group, which explains why both to the left and to the right of the cut-off, the treatment group is less likely to be subject to enforcement interventions. Importantly, the absence of discontinuities in the density and the pre-enforcement outcomes, both in the treatment and control group, seems to validate the use of a regression discontinuity design to estimate the causal effect of enforcement actions.

The impact of enforcement on compliance is illustrated in panel (b) of Figure 4. The fraction of taxpayers who have paid after 180 days is higher to the right than to the left of the threshold. Interestingly, compliance levels are similar in the treatment and control group to the right of the cut-off where enforcement intensity is high, while to the left where intensity is lower the treatment group is substantially more compliant.

To estimate the causal effects of the simplification treatments and the enforcement actions, we implement the standard regression discontinuity method in the control group, and add treatment dummies. Formally, let Y_i denote the tax compliance outcome of individual i , z_i their tax liability, c the tax liability cutoff. As before, S_i a dummy variable equal to one for the randomly assigned group who received the simplified letter and X_i is a vector of individual characteristics (see Table 1). The estimating equation is:

$$\begin{aligned} Y_i = & \alpha + \beta_S S_i + \beta_E 1\{z_i - c > 0\} + \beta_{S,E} S_i \times 1\{z_i - c > 0\} \\ & + \delta_{C,l}(z_i - c) + \delta_{C,r} 1\{z_i - c > 0\} \times (z_i - c) + \delta_{S,l} S_i \times (z_i - c) \\ & + \delta_{S,r} S_i \times 1\{z_i - c > 0\} \times (z_i - c) + \gamma X_i + \varepsilon_i \end{aligned}$$

Due to the random assignment, β_S is the effect of simplification at the cutoff from the left, where enforcement is weaker. Due to the regression-discontinuity, β_E identifies the effect of additional enforcement actions on tax compliance in the control group. Combining the two sources of variation, $\beta_{S,E}$ identifies the difference in treatment effects due to higher enforcement at the threshold. As in a typical regression discontinuity setting, $\delta_{C,l}$ and $\delta_{C,r}$ capture the relation between the forcing variable (tax liability) and the outcome (tax compliance) to the left and the right of the discontinuity, while $\delta_{S,l}$ and $\delta_{S,r}$ allow this relation to be different for the treatment group. An alternative interpretation is that the latter interaction terms allow for heterogeneity in treatment effects depending on the tax liability, both to the left and to the right of the cutoff.

Table 5 presents the corresponding regression results, using the Imbens-Kalyanaraman bandwidth computed for the control group in our experiment. We first consider the RDD estimates for the control group in our experiment. Column 1 confirms that the probability of enforcement increased by more than 50%, from 20 to 33%, at the threshold. Before

enforcement actions begun, the payment probability, however, was smooth at the threshold (Column 2). In contrast, 180 days after reminder receipt, the payment probability increased by 6pp at the threshold, reaching a probability of 88% for taxpayers in the control group to the right of the threshold (Column 3). Second, we consider the effects of simplification, not just on payment, but also on follow-up enforcement. As Column 1 shows, simplification decreased the probability of any enforcement action by almost half, from 20% in the control to 12%. This is due to the fact that simplified reminders made late payers 15pp more likely to pay before enforcement actions begun: from 49 to 64% (Column 2). Note that these effects are larger than those we report for the whole late payer sample (see Table 2). After 180 days, once payment rates in the control group have increased to 81%, the treatment effects were smaller, but still significant: a 4pp increase (Column 3). Finally, we estimate the difference in treatment effects to the left and to the right of the threshold. While the difference $\beta_{S,E}$ is not significant, the estimate is negative and large enough to mostly offset the positive treatment effect on the probability of paying at 180 days (Column 3).²¹ This confirms the graphical evidence that with high intensity enforcement the effects of simplification in the long run are virtually zero.

While the compliance benefits of nudges seem to disappear because of follow-up interventions on non-compliant taxpayers, they do bring important benefits by saving on enforcement costs as we discuss further below. Interestingly, we can also use our results to estimate the counterfactual effect of simplification after 180 days if the follow-up enforcement intervention had not taken place. Of course, in practice, the reminder letters effectiveness depends on tax payers' expectation of the follow-up enforcement by the administration. Still, to calculate the effect of simplification net of the crowd-out by the follow-up interventions, we impute the level of compliance based on the difference in compliance between high and low intensity enforcement groups scaled up by the difference in enforcement probability between them. Formally, let Y denote the payment probability, F the enforcement probability, z tax liability, c the cutoff and S letter simplification. Let the superscript F and Y denote the estimated coefficients when the dependent variable is F and Y , respectively. We approximate

²¹Note that these effects are driven by registered letters and garnishments (Appendix Table A.11).

the average treatment effect in absence of enforcement, ATE_0 , by:

$$\begin{aligned}
ATE_0 &\approx \left[E(Y|_{S=1,z<c}) - E(F|_{S=1,z<c}) \frac{E(Y|_{S=1,z>c}) - E(Y|_{S=1,z<c})}{E(F|_{S=1,z>c}) - E(F|_{S=1,z<c})} \right] \\
&- \left[E(Y|_{S=0,z<c}) - E(F|_{S=0,z<c}) \frac{E(Y|_{S=0,z>c}) - E(Y|_{S=0,z<c})}{E(F|_{S=0,z>c}) - E(F|_{S=0,z<c})} \right] \\
&= \left[\left(\widehat{\alpha^Y} + \widehat{\beta_S^Y} \right) - \left(\widehat{\alpha^F} + \widehat{\beta_S^F} \right) \frac{\left(\widehat{\beta_E^Y} + \widehat{\beta_{S,E}^Y} \right)}{\left(\widehat{\beta_E^F} + \widehat{\beta_{S,E}^F} \right)} \right] - \left[\widehat{\alpha^Y} - \widehat{\alpha^F} \frac{\widehat{\beta_E^Y}}{\widehat{\beta_E^F}} \right] = 0.09
\end{aligned}$$

This calculation also relies on a homogeneity assumption: we need that the effect of enforcement on the payment probability is the same for taxpayers who pay only when enforcement intensity increases from below to above the threshold and for taxpayers who pay even with low intensity enforcement. The counterfactual analysis suggests that in absence of the follow-up enforcement actions, the effect of simplification on the payment probability of late payers would have been 9pp after 180 days, which is approximately two-thirds of the effect estimated before enforcement actions begun (15pp).

5.2 Cost-Effectiveness and Welfare

We evaluate the cost-effectiveness of the simplification treatment in three different ways. First, we compare the benefits of the treatment in terms of additional revenue and savings on enforcement actions to the costs of simplifying the tax correspondence. Second, we compare the cost of raising one euro of extra revenue through reminder simplification and through enforcement actions. Finally, we calculate the total cost of enforcement actions that would be needed to raise the same extra revenue as the simplification treatment, using the counterfactual effect of simplification in the absence of follow-up enforcement.

The first method is based on experimental results only. To compute extra revenues, we estimate the effect of simplified letters on the probability of paying taxes as late as possible in the tax cycle, which is 180 days after the payment deadline, and assume that after this date the treatment effect will remain constant.²² As Appendix Table A.9 shows, the estimated treatment effect on the probability of payment at 180 days is about 1pp, which we multiply by the average amount paid, conditional on a payment, at that date (€1,615) and the number of tax payers in the treatment group (204,223) to obtain total extra revenues equal to €3.14 million. To compute savings on the cost of enforcement, we estimate the effect of simplified letters on the probability of the three most common forms of enforcement

²²After 180 days, tax filing for the next fiscal year begins: the administrative data that we use does not allow us to track outstanding debts separately from new tax liabilities.

actions – registered letters, garnishment and bailiffs. Multiplied by the cost of the respective enforcement measures, we obtain a total cost saving of €0.70 million.²³ Adding the extra revenues and costs savings on enforcement, the total benefit of the intervention equals €3.84 million. In comparison, the costs of simplification were negligible: the administration paid €69,300 for the design of the new letter, including ICT staff, data analysts, legal experts, communication staff and management, and the printing of the new (colored) letter costs an extra €0.05 per letter. The total cost of simplifying the reminder letters amounts to €79,511 and is about 50 times smaller than its benefits. Simplifying the reminder letters was thus a high return investment for the tax administration.

The second method builds on the regression discontinuity results from the previous section. Since we are able to estimate the compliance effects of the simplification treatment and the enforcement interventions separately, we can ask what the most cost-effective way is to raise one euro of extra revenue. It is well known in optimal tax theory that from an efficiency perspective the excess burden of different taxes should be equalized at the margin. Extending this insight, [Keen and Slemrod \(2017\)](#) have shown that in absence of equity considerations the excess burden of each tax should be equalized with the marginal cost of interventions to enforce payment of the tax. For the enforcement interventions, we first use regression discontinuity estimates for the increase in the probability that registered letters (10.2pp) and garnishment (5.9pp) are sent at the threshold (see Appendix Table [A.11](#)) and their cost (€5.7 and €17.1 respectively) to compute the cost of the increase in enforcement intensity at the threshold, which is €1.6.²⁴ We then use regression discontinuity estimates of the effect of enforcement intensity on the probability of payment at 180 days (from Table [5](#)) multiplied by average payments made at the threshold to estimate additional revenues raised. The ratio of the two, i.e., the cost of raising one more euro of tax revenues through enforcement is equal to €0.27. This estimate is arguably in the range of standard estimates of the marginal excess burden of personal income taxes, suggesting that the enforcement intensity may well be desirable ([Keen and Slemrod, 2017](#)). In comparison, the resource cost of using nudge interventions is much smaller. As explained above, the cost of simplification was €79,511, or €0.39 per letter sent. We multiply our counterfactual estimate of the effect of simplification on the probability of payment in the absence of follow-up enforcement

²³As Appendix Table [A.10](#) shows, the estimated treatment effects on follow-up enforcement are –7.4pp for registered letters, –2.8pp for garnishment actions and –1.2pp for bailiffs. Multiplying these figures by the cost of each action and the number of treated taxpayers, we obtain costs savings of $€5.7 * 0.074 * 204,223 = 86,141$ for registered letters, $€17.1 * 0.028 * 204,223 = 97,084$ for garnishment and $€213 * 0.012 * 204,223 = 513,294$ for bailiffs.

²⁴As Appendix Table [A.11](#) shows, there is no significant increase in the use of bailiff at the threshold. As an enforcement tool, the use of bailiffs is applied to debts of relatively large amounts, while registered letters and garnishments are more often employed.

by the average tax payment, and obtain €8.74 extra revenue per letter. Hence the cost of raising one euro with simplified reminders is €0.05, which is six times smaller than with enforcement actions.²⁵ This second method confirms that simplifying reminders is far more cost-effective than intensifying enforcement.

The third method extrapolates the regression discontinuity results to the whole sample, using a back-of-the envelope calculation. At the enforcement threshold, the treatment effect was 15pp after 14 days and the counterfactual effect absent follow-up enforcement at 180 days was 9pp (Table 5). Hence for the whole sample the estimated treatment effect of 10.3pp after 14 days suggests that the counterfactual effect, in the absence of follow-up enforcement, would have been $10.3 * 9/15 = 6$ pp at 180 days. Multiplying this figure by taxes paid by the treatment group gives €20.2million of extra revenue. To obtain these extra revenues with traditional enforcement methods at the cost of 27 cents per euro raised, the government would have had to spend €5.4 million. This is again substantially higher than the cost of the simplification intervention.

Regardless of the method we use for the cost-benefit analysis, simplifying letters seems highly cost effective, in itself and when compared to the alternative of using standard enforcement actions. The above calculations, however, ignore other welfare-relevant considerations that may be important when assessing the use of nudges. First of all, the letter treatments - when successful - changed the net transfers between taxpayers and the government, not only by affecting the taxes paid, but also avoiding the late penalties and interests on outstanding tax liability. Second, the nudges can affect individuals' welfare above and beyond their after-tax income. The simplified correspondence reduces compliance costs, but may also reduce the disutility of paying taxes. While the same may be true for highlighting the public value of taxes paid, the opposite effect seems as plausible when using deterrence or invoking social norms.²⁶ Finally, the nudges may have heterogeneous effects and differentially affect different groups, which may enter the welfare considerations. We turn to these heterogeneous effects next.

5.3 Heterogeneous Effects

Understanding heterogeneity in treatment effects is important for assessing the potential distributional welfare consequences of an intervention, which will also depend on the incidence

²⁵We consider this a conservative estimate as the cost of nudging is largely driven by the fixed costs of experimental design. If these are ignored the per letter cost goes down to 0.05 making it eight times cheaper and thus lowering significantly the cost to benefit ratio of the nudging intervention.

²⁶For example, Di Tella et al. (2015) show that complexity can lead people to be “conveniently upset” and use it as an excuse not to comply.

of frictions that the intervention tries to overcome.²⁷ Understanding the heterogeneity can also be instrumental for improving the targeting of interventions, just like the tax administration targets their enforcement actions to different sub-populations depending on their cost-effectiveness. We focus again on the payment reminder experiments, for which we were able to obtain a large set of observables (including various demographics like age, family composition, region, amount owed, taxable income and solvency estimated by the tax administration). To discipline our analysis of treatment effect heterogeneity, we use the causal forests algorithm created by [Wager and Athey \(2018\)](#).²⁸

Figure 6 plots the dispersion of the treatment effects by treatment category (bin size is set to 0.005 for all figures). While the figure only uncovers the heterogeneity in treatment effects based on observables, it is interesting to compare the predicted heterogeneity across treatments, using the same observables. Indeed, we see a wide dispersion for the simplification treatment, but less so for the deterrence and tax morale ones. Moreover, perhaps not surprisingly, the effect of the simplification treatment never turns negative. The deterrence treatment, however, has negative effects for some tax payers. Interestingly, the tax morale treatments seem to backfire across the population: almost all estimated treatment effects are negative.

Using the same causal forests estimates, we can determine which observable characteristics drive the heterogeneity in treatment effects. Figures A.6a to A.6h present the average of the different observables in each treatment effect quintile. Figure A.6a suggests that simplification is the least effective for on average older taxpayers and deterrence is the most effective for on average younger taxpayers. Simplification seems to be very effective among taxpayers with children, while the tax morale treatments seems to be more likely to backfire for this group (Figure A.6c). Simplification seems also to be particularly ineffective for tax payers with high predicted solvency, as predicted by the tax administration (Figure A.6f). Deterrence, finally, seems to be particularly effective for taxpayers with lower outstanding tax liability (Figure A.6h). There is no obvious pattern for gender (Figure A.6b), language (Figure A.6d), region (Figure A.6e) or income (Figure A.6g).

The machine learning results thus identify four relevant dimensions of treatment heterogeneity: age, number of children, tax liability and solvency. We next estimate OLS regressions of tax compliance on interactions of the treatment with these four main characteristics, while including interactions of the treatment with all other characteristics as controls. Table 6 presents the estimates. The corresponding estimates for the second TPR experiment,

²⁷See for example [Alcott et al. \(2018\)](#) in the context of using corrective sin taxes.

²⁸According to [Chernozhukov et al. \(2018\)](#), we are in the case where the [Wager and Athey \(2018\)](#) method provides robust results: we have 10 dimensions of heterogeneity and about 230,000 observations ($\log(230,000) = 12 > 10$).

which was implemented in the fiscal year 2015 and was not used in the prediction algorithm, are reported in Appendix Table [A.12](#).

The first main dimension of heterogeneity detected by the causal forests is age. Once we control for all observable characteristics and their interaction with treatment, we find no differential treatment effect of simplification depending on taxpayers' age. However, we do find evidence of a significantly lower effect of both the deterrence and the tax morale treatments for older workers. As compared to the younger group (30 and below), taxpayers aged 40 and above have between 2 and 3 pp lower treatment effects. The fact that these groups in the control have much higher compliance levels is likely part of the explanation as there is a lower margin for improvement. But it could also be that nudge messages are less effective with an older public.

The second main dimension of heterogeneity is the number of children. Households with more children are significantly less compliant in the control, with a 2-3pp lower likelihood of paying before 14 days than the control mean of 40%. The simplified letter, however, increases the compliance by taxpayers with one child by an additional 1.9 pp (se 1.3) and with two or more children by an additional 2.7 pp (se 1.4). Households with children are hence less compliant on average, but reducing the information overload seems particularly helpful for them.

The third important dimension of heterogeneity is outstanding tax liability. The compliance effects of the simplified letter are about 4.2 to 6.2pp smaller for taxpayers with outstanding debt above the lowest quintile. This decrease is replicated in 2016 and remains when including all other controls interacted with the treatment variables. This evidence indicates a trade-off in targeting taxpayers with higher outstanding tax liability. While the returns to making them compliant are larger, their tax compliance is lower overall and so is their responsiveness to the simplified letter treatment. The deterrence treatments also have a significantly smaller and even negative effect on the higher quintiles of taxpayers in terms of outstanding debt. However, this may also be driven by the fact that it is the average penalty that is made explicit, rather than an individual-specific projection of the potential penalty.

Finally, the fourth dimension of treatment heterogeneity is the solvency score. This score is computed by the tax administration at the start of the tax collection cycle using a prediction model and then used to decide on follow-up enforcement actions. The baseline compliance rate ranges from 20.9 percent for the lowest quintile to 73.1 percent for the highest quintile, confirming the strong predictive value of the solvency score in this context. Interestingly, we find a non-monotone relation in the treatment effect. The simplification treatment increases compliance the least in the lowest and the highest quintile and the most

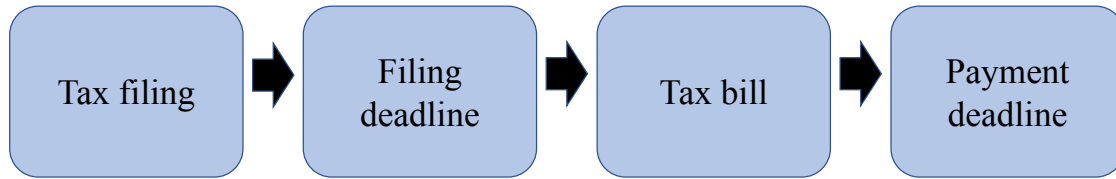
in the middle quintiles. The differences in treatment effects are substantial, with a maximum difference in treatment effects of 6pp. These results highlight that simply targeting further interventions based on baseline compliance may be undesirable.

6 Conclusion

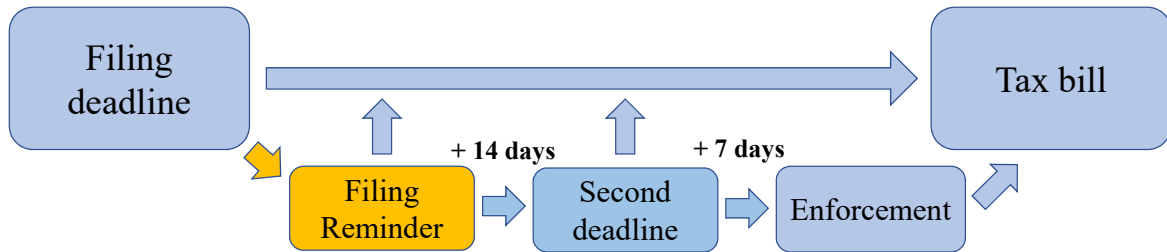
By way of a series of population-wide experiments in Belgium, this research has shown that simplifying communication by the tax administration consistently improves tax compliance. Simplification makes taxpayers pay taxes on time and makes both late filers and late payers comply more swiftly than they would otherwise. The positive effects of simplification are universal across the population, they are sustained when simplification is repeated, and they persist in the next fiscal years. Making it as easy as possible to comply therefore deserves even greater attention since communication is an inherent part of any tax administration. Our experimental design allows us to compare simplification with deterrence and tax morale treatments in the same setting thereby simultaneously testing the three main factors of tax compliance studied in the literature. The results also demonstrate the effectiveness of deterrence messages. In contrast, invoking tax morale does not raise compliance and even backfires for most taxpayers. Finally, we are able to causally estimate the costs and benefits of nudge-type interventions as compared to traditional enforcement levers, and find nudge treatments to be highly cost effective.

Graphs

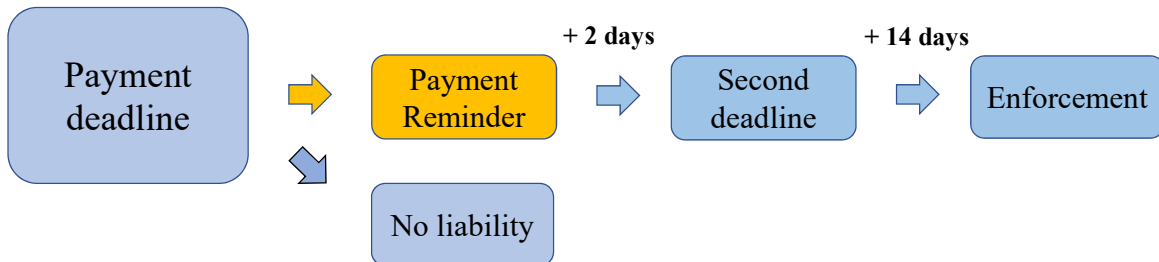
Figure 1: Tax process



(a) Filing and payment

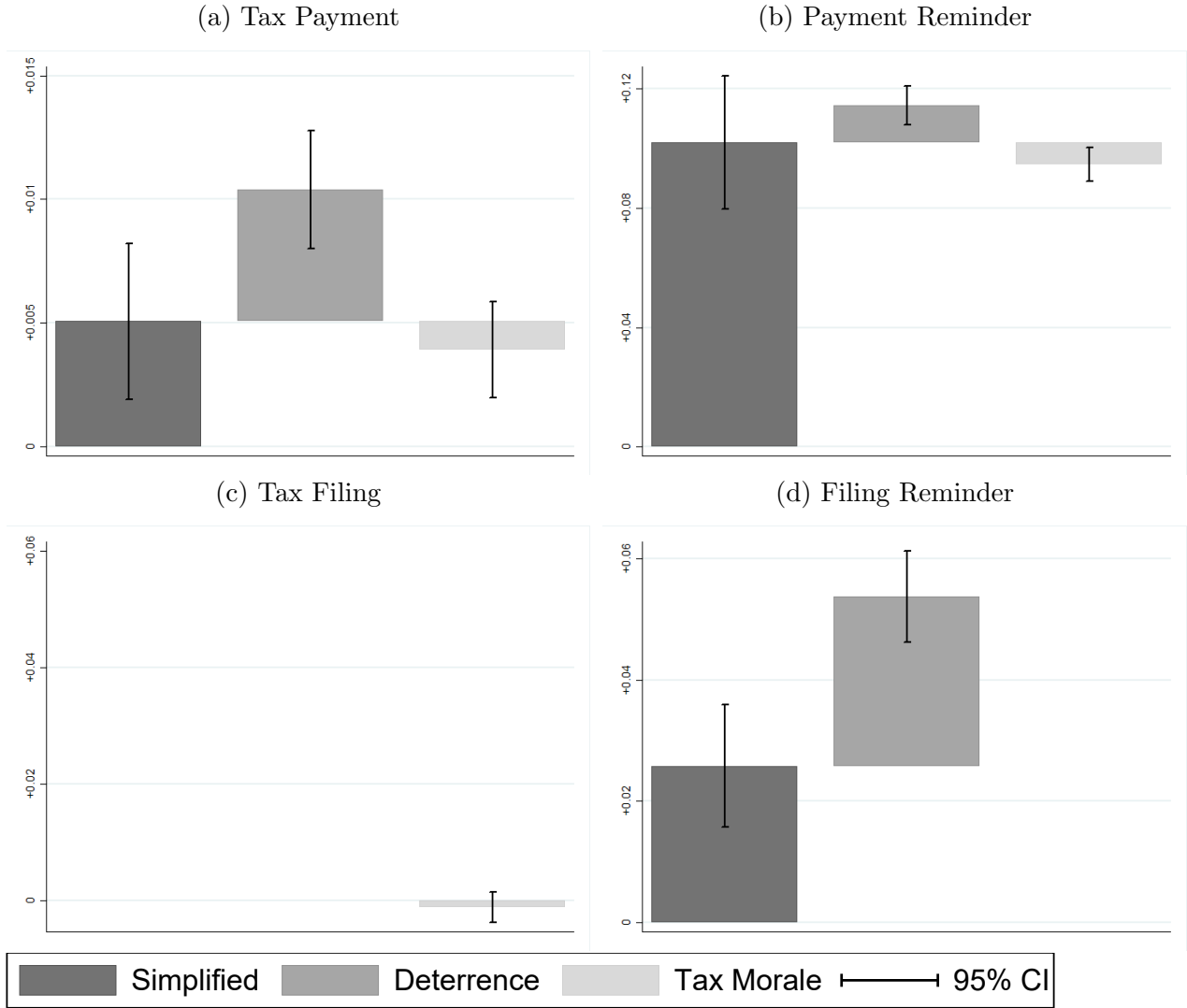


(b) Filing reminder process



(c) Payment reminder process

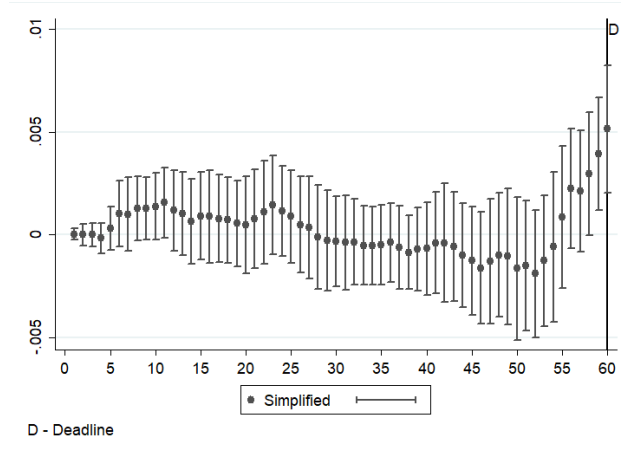
Figure 2: Summary of the Main Results



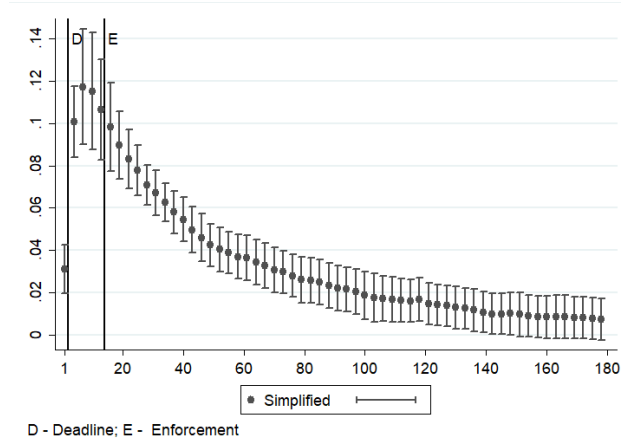
Note: The figure presents treatment effect estimates from baseline specifications for the TP (Panel (a)), TPR FY2014 (Panel (b)), TF (Panel (c)) and TFR FY2015 (Panel (d)) experiments. The outcome is partial payment probability at 60 days (deadline) in Panel (a), and at 14 days (enforcement) in Panel (b). The outcome is reported taxable income in Panel (c) and filing probability at 21 days (enforcement) in Panel (d). Control variables are listed in Table 1, for exact estimates refer to Table 2. 95% confidence intervals based on robust standard errors are plotted. Standard errors are clustered by date of letter receipt in Panels (a) and (b).

Figure 3: Dynamic Effects of Simplification

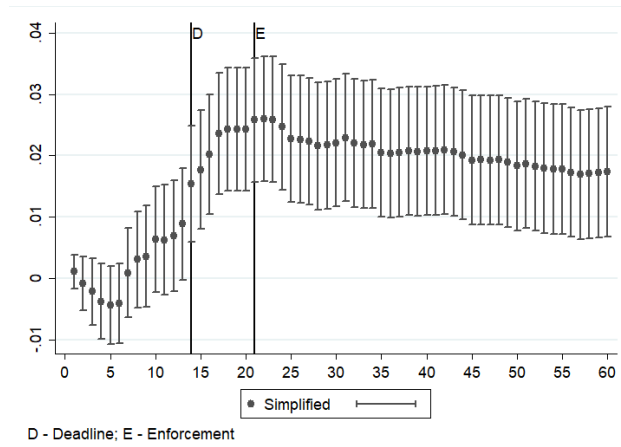
(a) Tax Payment



(b) Payment Reminder



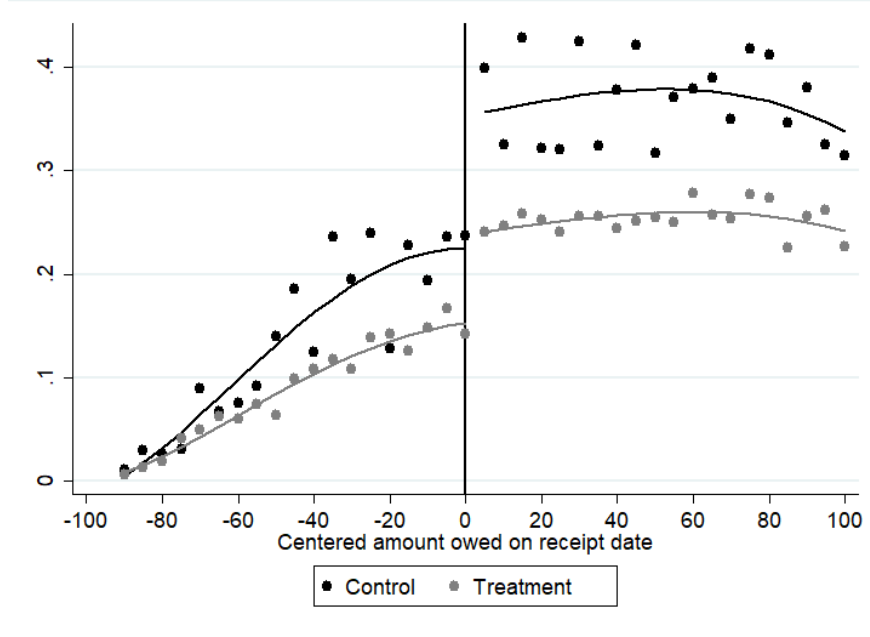
(c) Filing Reminder



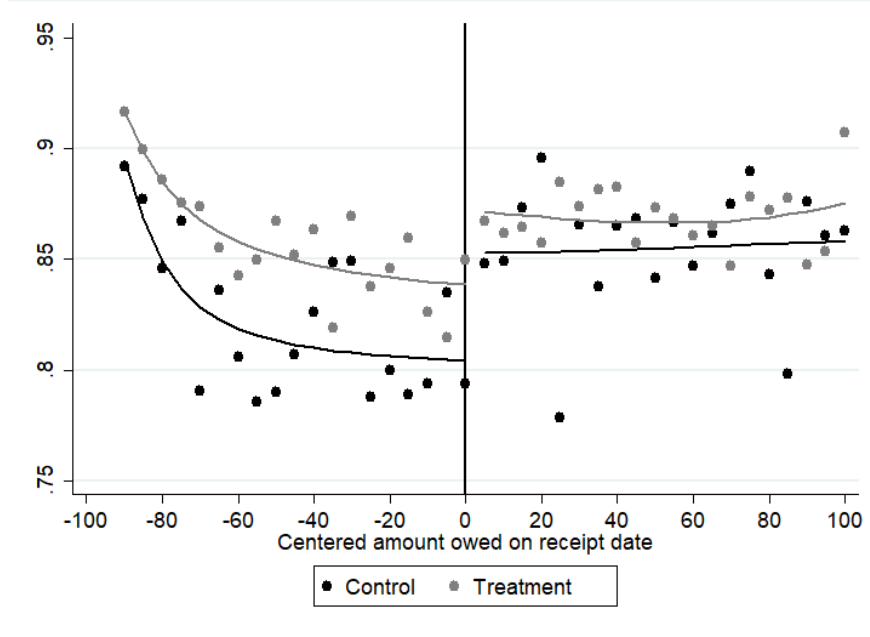
Note: The figure presents simplification treatment effect estimates by days since letter receipt for the TP (Panel (a)), TPR FY2014 (Panel (b)) and TFR FY2015 (Panel (c)) experiments. The outcome is partial payment probability in Panels (a) and (b), and filing probability in Panel (c). The vertical lines indicate the payment/filing deadline and/or the day enforcement actions start. Control variables are listed in Table 1. 95% confidence intervals based on robust standard errors plotted. Standard errors are clustered by date of letter receipt in Panels (a) and (b).

Figure 4: Effects of Enforcement and Simplification

(a) Probability of Enforcement after 180 Days

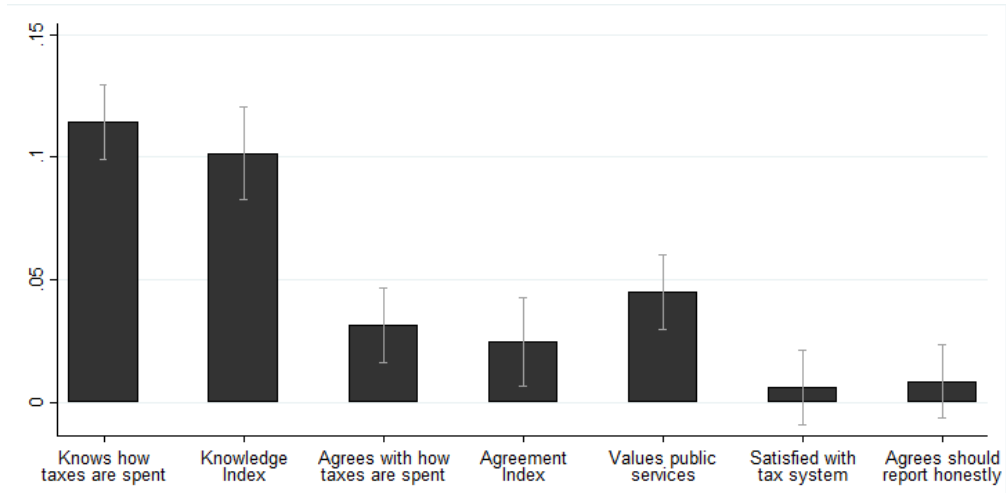


(b) Probability of Partial Payment after 180 Days



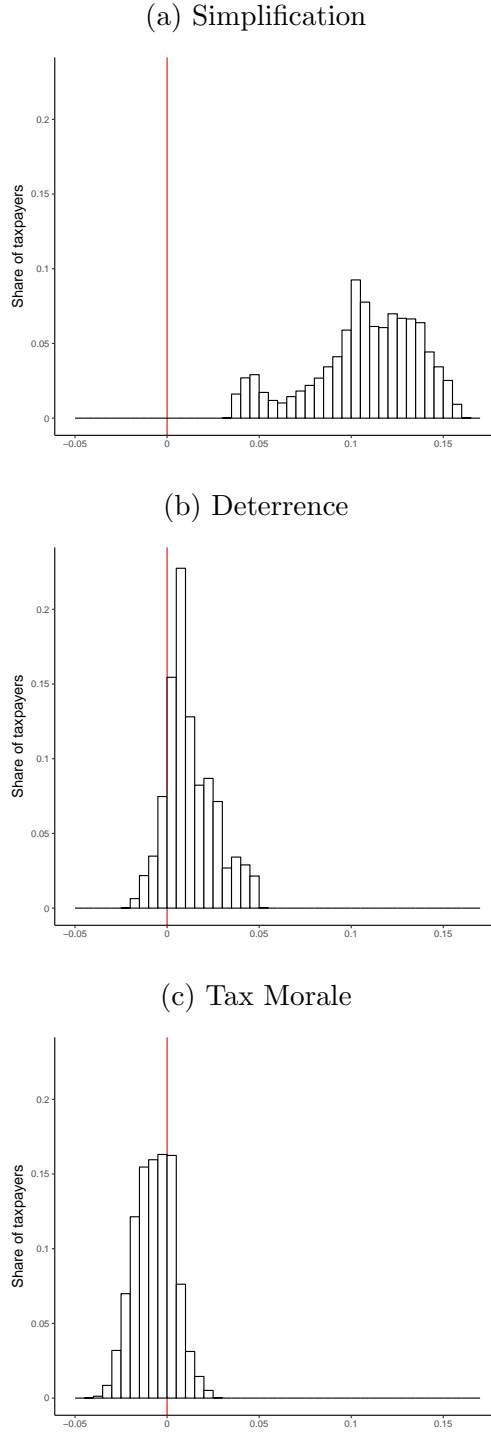
Note: The figure is based on the TPR FY2014 experiment. It shows probability of enforcement after 180 days (Panel (a)) and probability of partial payment after 180 days (Panel (b)) by initial amount owed (centred at the enforcement threshold). Bin size is set to €5 and amounts within €100 of the enforcement threshold are considered. Fractional polynomial predictions plotted.

Figure 5: Tax Filing: Treatment Effect on Survey Responses



Note: The figure presents treatment effect estimates from the analysis of survey responses in the TF experiment. Outcomes are standardized using mean and standard deviation of control group responses. Control variables are dummies for gender and age categories. 95% confidence intervals based on robust standard errors are plotted.

Figure 6: Distribution of treatment effects



Note: The figure presents the distribution of estimated treatment effects in the TPR FY2014 experiment. It uses the generalized random forest (GRF) algorithm ([Wager and Athey, 2018](#)) as described in the text. Figures (a)-(c) differ in the definition of treatment and control groups. In Figure (a) the control is composed of taxpayers who received the old letter and the treatment of taxpayers who received a simplified letter without any additional message. In Figure (b) and (c) taxpayers who received a simplified letter without any additional message are the control group. In Figure (b) the treatment is composed of taxpayers who received a simplified letter with a deterrence message. In Figure (c) the treatment is composed of taxpayers who received a simplified letter with an added tax morale message.

Tables

Table 1: Summary Statistics of Control Variables

Experiment:	All taxpayers	Tax Payment	Payment Reminder	Tax Filing	Filing Reminder
	(1)	(2)	(3)	(4)	(5)
<i>Demographics</i>					
Male dummy	0.309 (0.462)	0.324 (0.468)	0.448 (0.497)	0.276 (0.447)	0.529 (0.499)
Couple dummy	0.346 (0.476)	0.415 (0.493)	0.298 (0.457)	0.445 (0.497)	0.132 (0.339)
Age	49.495 (18.129)	53.354 (16.382)	47.764 (15.611)	47.596 (15.585)	42.229 (16.249)
Number of children	0.413 (0.869)	0.351 (0.771)	0.409 (0.830)	0.579 (0.950)	0.334 (0.836)
Married dummy				0.476 (0.499)	
Widowed dummy				0.040 (0.196)	
Divorced dummy				0.156 (0.363)	
<i>Region / Language</i>					
Wallonia dummy	0.327 (0.469)	0.316 (0.465)	0.367 (0.482)	0.284 (0.451)	0.390 (0.488)
Flanders dummy	0.570 (0.495)	0.596 (0.491)	0.525 (0.499)	0.637 (0.481)	0.390 (0.488)
French dummy	0.421 (0.494)	0.386 (0.487)	0.473 (0.499)	0.357 (0.479)	0.592 (0.491)
German dummy	0.006 (0.076)	0.011 (0.104)	-	0.003 (0.051)	0.007 (0.084)
<i>Other</i>					
Amount owed	568.635 (7301.068)	2676.205 (11869.230)	1890.950 (4746.221)		
Income			33211.010 (28804.210)		
Solvency score			11.657 (4.674)		
N	6,689,779	1,216,317	229,751	942,571	148,925

Note: The table presents means and standard deviations (in parentheses) of control variables for different samples. In column 1 the sample is composed of all individual income taxpayers in FY2016. In column 2 it is the sample of the TP FY2016 experiment. In column 3 it is the sample of the TPR FY2014 experiment. In column 4 it is the sample of the TF FY2016 experiment. In column 5 it is the sample of the TFR FY2015 experiment. The base category for gender is female, for region Brussels, for language Flemish and for marital status single.

Table 2: Main Results

Panel A: Payment	Probability of some payment	
	at 60 days (deadline)	at 14 days (before enforcement)
	Tax Payment (1)	Payment Reminders (2)
Simplified	0.00507*** (0.00146)	0.102*** (0.0103)
+ Deterrence	0.00533*** (0.00110)	0.0124*** (0.00303)
+ Tax Morale	-0.00116 (0.000899)	-0.00742** (0.00259)
P-value (Det=TM)	0.000	0.000
Control mean	0.728	0.447
N	1,216,317	229,751
Panel B: Filing	Probability of having filed	
	Log pre-check taxable income	at 21 days (before enforcement)
	Tax Filing (1)	Filing Reminders (2)
Simplified		0.0258*** (0.00516)
+ Deterrence		0.0279*** (0.00385)
Tax Morale	-0.00110 (0.00135)	
Control mean	15.04	0.317
N	942,571	148,925

Note: The table presents treatment effect estimates from baseline specifications in four separate experiments. Column 1 in Panel A presents the results of the TP experiment (taxpayers for the FY2016). Column 2 in Panel A presents the results of the TPR 2014 experiment (late taxpayers in the FY2014). Column 1 in Panel B presents the results of the TF experiment (online tax filers in the FY2016). Column 2 in Panel B presents the results of the TFR experiment (late tax filers in the FY2015). Control variables are listed in Table 1. Robust standard errors in parentheses, clustered by date of letter receipt in Panel A. * p<0.1; ** p<0.05; *** p<0.01.

Table 3: Replication, Long-term and Repeated Treatment Effects

Panel A: Long-term Effects	Probability of being on time with payment FY+1 year (1)	Probability of being on time with payment FY+2 years (2)
Simplified	0.0114*** (0.00276)	0.00554 (0.00378)
+ Deterrence	-0.000929 (0.00250)	-0.00378 (0.00214)
+ Tax Morale	-0.00728*** (0.00229)	-0.00606** (0.00238)
P-value (Det=TM)	0.001	0.261
Control mean	0.689	0.769
N	229,751	229,751
Panel B: Replication FY2015	Probability of some payment at 14 days (before enforcement)	
Simplified	0.104*** (0.00479)	
+ Deterrence	0.0160*** (0.00294)	
+ Tax Morale	-0.0113*** (0.00290)	
+ Deterrence & Tax Morale	0.00685* (0.00327)	
P-value (Det=TM)	0.000	
P-value (Det=Det+TM)	0.003	
Control mean	0.419	
N	188,180	
Panel C: Repeated Treatment	Probability of some payment at 14 days (before enforcement)	
Simplified 2014	-0.00402 (0.0117)	
Simplified 2015	0.0966*** (0.0122)	
Simplified 2014 * Simplified 2015	0.000960 (0.0121)	
Control mean	0.423	
N	64,736	

Note: The table presents results from the replication, long-term and repeated treatment analysis. The sample in Panel A is the universe of late payers in FY2015. In Panel B it is the universe of late payers in FY2014. In Panel C it is composed of taxpayers who were late with payment in both FY2014 and FY2015. Control variables are listed in Table 1. Standard errors in parentheses are clustered by date of letter receipt.
* p<0.1; ** p<0.05; *** p<0.01.

Table 4: Tax Filing: Survey Results

	Satisfied with tax system (1)	Values public services (2)	Agrees should be honest (3)	Knows how taxes are spent (4)	Knowledge index (5)	Agrees with how taxes are spent (6)	Agreement index (7)
Treatment	0.00615 (0.00773)	0.0449*** (0.00773)	0.00880 (0.00770)	0.114*** (0.00765)	0.101*** (0.00963)	0.0317*** (0.00774)	0.0247*** (0.00923)
N	66,874	66,698	66,607	66,530	47,194	66,430	47,897

Note: The table presents treatment effect estimates from the analysis of survey responses in the tax filing experiment (TF FY2016). Outcomes are standardized using mean and standard deviation of control group responses. Control variables are dummies for gender and age categories. Robust standard errors in parentheses. * $p < 0.01$; ** $p < 0.05$; *** $p < 0.01$.

Table 5: RDD: Effect of Payment Reminders Simplification and Enforcement FY2014

	Probability of enforcement within 180 days (1)	Probability of some payment at 14 days (before enforcement) (2)	within 180 days (3)
Simplified	-0.0829*** (0.0217)	0.151*** (0.0249)	0.0442** (0.0194)
Enforcement	0.130*** (0.0296)	0.00626 (0.0340)	0.0615** (0.0265)
Simplified*Enforcement	-0.0482 (0.0316)	0.000312 (0.0363)	-0.0273 (0.0283)
Control Mean	0.200	0.489	0.813
N	20,793	23,312	21,894

Note: The table presents regression discontinuity design estimates. Simplified is a dummy variable equal to one for taxpayers who received a simplified letter. Enforcement is a dummy variable equal to one for liability amounts above the cut-off value. Control variables are listed in Table 1. Standard errors in parentheses are clustered by date of letter receipt. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 6: Heterogeneous Effects – Payment Reminder Experiment FY2014

Probability of payment at 14 days (before enforcement)			
Variable	(1)	Variable	(2)
Simplified	0.112*** (0.0139)	Solvency score Q2 * Simplified	0.0586*** (0.0105)
+ Deterrence	0.0529*** (0.00908)	* Deterrence	-0.00793 (0.00710)
+ Social	0.0136 (0.0177)	* Social	-0.0126** (0.00536)
Age 31-40 * Simplified	0.0123 (0.0128)	Solvency score Q3 * Simplified	0.0564*** (0.0142)
* Deterrence	-0.0105 (0.0118)	* Deterrence	-0.00357 (0.00775)
* Social	-0.00637 (0.0103)	* Social	-0.00428 (0.0105)
Age 41-50 * Simplified	0.0246 (0.0139)	Solvency score Q4 * Simplified	0.0239 (0.0155)
* Deterrence	-0.0260** (0.0117)	* Deterrence	-0.0162 (0.0132)
* Social	-0.0258** (0.0115)	* Social	-0.00116 (0.0147)
Age 51-60 * Simplified	0.0113 (0.0131)	Solvency score Q5 * Simplified	-0.0303 (0.0208)
* Deterrence	-0.0284** (0.0102)	* Deterrence	-0.00214 (0.0110)
* Social	-0.0280** (0.00993)	* Social	0.0115 (0.0115)
Age 61+ * Simplified	-0.0167 (0.0128)	Liability Q2 * Simplified	-0.0495*** (0.0118)
* Deterrence	-0.0245* (0.0131)	* Deterrence	-0.00485 (0.0102)
* Social	-0.0165* (0.00818)	* Social	0.0174 (0.0109)
One child * Simplified	0.0191 (0.0128)	Liability Q3 * Simplified	-0.0423*** (0.0102)
* Deterrence	0.00770 (0.0104)	* Deterrence	-0.0185** (0.00722)
* Social	0.0133 (0.0118)	* Social	0.00357 (0.00730)
Two or more children * Simplified	0.0269* (0.0141)	Liability Q4 * Simplified	-0.0622*** (0.0102)
* Deterrence	-0.0110 (0.0116)	* Deterrence	-0.0161 (0.00962)
* Social	-0.0116 (0.0106)	* Social	0.0163 (0.00990)
		Liability Q5 * Simplified	-0.0456*** (0.0112)
		* Deterrence	-0.0410*** (0.00783)
		* Social	0.00652 (0.00989)
Control mean			0.400
N			229,751

Note: The table presents treatment effect estimates from the heterogeneous effects analysis. Control variables are listed in Table 1. The full set of interactions between individual control and treatment variables are included in the estimation (coefficients not reported). Standard errors in parentheses are clustered by date of letter receipt. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

References

- Abeler, J. and S. Jäger (2015). Complex tax incentives. *American Economic Journal: Economic Policy* 7(3), 1–28.
- Aghion, P., U. Akcigit, M. Lequien, and S. Stantcheva (2017). Tax simplicity and heterogeneous learning. Technical report, National Bureau of Economic Research.
- Alcott, H., B. Lockwood, and D. Taubinsky (2018). Regressive sin taxes, with an application to the optimal soda tax. Working paper.
- Allcott, H. and T. Rogers (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review* 104(10), 3003–37.
- Allingham, M. and A. Sandmo (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics* 1, 323–338.
- Alm, J. (2012). Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies. *International Tax and Public Finance* 19(1), 54–77.
- Alstadsæter, A., N. Johannesen, and G. Zucman (2018). Who owns the wealth in tax havens? macro evidence and implications for global inequality. *Journal of Public Economics* 162, 89–100.
- Andreoni, J., B. Erard, and J. Feinstein (1998). Tax compliance. *Journal of Economic Literature* 36, 818–860.
- Benartzi, S., J. Beshears, K. L. Milkman, C. R. Sunstein, R. H. Thaler, M. Shankar, W. Tucker-Ray, W. J. Congdon, and S. Galing (2017). Should governments invest more in nudging? *Psychological science* 28(8), 1041–1055.
- Benzarti, Y. (2017). How taxing is tax filing? using revealed preferences to estimate compliance costs. Working Paper 23903, National Bureau of Economic Research.
- Besley, T., A. Jensen, and T. Persson (2019). Norms, enforcement, and tax evasion. Working Paper 25575, National Bureau of Economic Research.
- Bhargava, S. and D. Manoli (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an irs field experiment. *American Economic Review* 105(11), 3489–3529.

- Blumenthal, M., C. Christian, J. Slemrod, and M. G. Smith (2001). Do normative appeals affect tax compliance? evidence from a controlled experiment in minnesota. *National Tax Journal*, 125–138.
- Bott, K. M., A. W. Cappelen, E. Sorensen, and B. Tungodden (2017). You’ve got mail: a randomised field experiment on tax evasion. Department of Economics, Norwegian School of Economics Discussion Paper 26/2014.
- Brockmeyer, A., M. Hernandez, S. Kettle, and S. Smith (2019). Casting a wider tax net: Experimental evidence from costa rica. *American Economic Journal: Economic Policy*.
- Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018). Generic machine learning inference on heterogenous treatment effects in randomized experiments. NBER Working Paper 24678, National Bureau of Economic Research.
- Chetty, R. and E. Saez (2013). Teaching the tax code: Earnings responses to an experiment with eitc recipients. *American Economic Journal: Applied Economics* 5(1), 1–31.
- Cox, J. C., D. Kreisman, and S. Dynarski (2018). Designed to fail: Effects of the default option and information complexity on student loan repayment. Working Paper 25258, National Bureau of Economic Research.
- Cranor, T., J. Goldin, T. Homonoff, and L. Moore (2018). Communicating tax penalties to delinquent taxpayers: Evidence from a field experiment. Working paper.
- Cronqvist, H., R. H. Thaler, and F. Yu (2018). When nudges are forever: Inertia in the swedish premium pension plan. In *AEA Papers and Proceedings*, Volume 108, pp. 153–58.
- Del Carpio, L. (2014). Are the neighbors cheating? evidence from a social norm experiment on property taxes in peru. Economics Department, Princeton University.
- Di Tella, R., R. Perez-Truglia, A. Babino, and M. Sigman (2015). Conveniently upset: Avoiding altruism by distorting beliefs about others’ altruism. *American Economic Review* 105(11), 3416–42.
- Guyton, J., D. S. Manoli, B. Schafer, and M. Sebastiani (2016). Reminders & recidivism: evidence from tax filing & eitc participation among low-income nonfilers. Working Paper 21904, National Bureau of Economic Research.
- Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev (2015). The making of homo honoratus: from omission to commission. Working Paper 21210, National Bureau of Economic Research.

- 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.
- HM Revenue & Customs (2018). Measuring tax gaps 2018 edition. Technical report, HM Revenue & Customs.
- Hoopes, J. L., D. H. Reck, and J. Slemrod (2015). Taxpayer search for information: Implications for rational attention. *American Economic Journal: Economic Policy* 7(3), 177–208.
- Internal Revenue Service (2016). Tax gap estimates for tax years 2008–2010. Technical report, Internal Revenue Service.
- Jensen, A. (2019). Employment structure and the rise of the modern tax system. Working Paper 25502, National Bureau of Economic Research.
- John, P. and T. Blume (2018). How best to nudge taxpayers? the impact of message simplification and descriptive social norms on payment rates in a central london local authority. *Journal of Behavioral Public Administration* 1(1).
- Keen, M. and J. Slemrod (2017). Optimal tax administration. *Journal of Public Economics* 152, 133–142.
- Kirchler, E. (2007). *The economic psychology of tax behaviour*. New York, NY: Cambridge University Press.
- Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica* 79(3), 651–692.
- Kleven, H. J. and W. Kopczuk (2011). Transfer program complexity and the take-up of social benefits. *American Economic Journal: Economic Policy* 3(1), 54–90.
- Kleven, H. J., C. T. Kreiner, and E. Saez (2016). Why can modern governments tax so much? an agency model of firms as fiscal intermediaries. *Economica* 83(330), 219–246.
- Luttmer, E. F. and M. Singhal (2014). Tax morale. *Journal of Economic Perspectives* 28(4), 149–168.
- Naritomi, J. (2018). Consumers as tax auditors. *CEPR Discussion Paper No. DP13276*.

- OECD (2010). Understanding and influencing taxpayers' compliance behavior. Technical report, OECD Centre for Tax Policy and Administration.
- 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. *American Economic Review* 105(8), 2539–69.
- Slemrod, J. (2018). Tax compliance and enforcement. Working Paper 24799, National Bureau of Economic Research.
- 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.
- Slemrod, J. and S. Yitzhaki (2002). Tax avoidance, evasion, and administration. In A. J. Auerbach and M. Feldstein (Eds.), *Handbook of Public Economics*, Volume 3 of *Handbook of Public Economics*, Chapter 22, pp. 1423–1470. Elsevier.
- Torgler, B. (2007). *Tax compliance and tax morale: A theoretical and empirical analysis*. Northampton, MA: Edward Elgar Publishing.
- Wager, S. and S. Athey (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association* 113(523), 1228–1242.

ONLINE APPENDIX

A.1 Filing Reminder Experiment 2014

For the fiscal year 2014, the administration independently carried out an experiment on filing reminders. 162,682 late filers were part of the experiment. Unlike the other experiments we study in the paper, the treatment group did not receive a simplified letter. Instead, they received tax morale messages included in the usual (complex) letter. The messages were as follows:

- Public Goods: “By filing and paying your taxes you guarantee the provision of essential services of the government, such as education, public health and public safety”
- Social Norm: “You belong to a minority because 94% of Belgian taxpayers filed on time. Why not follow their example?”
- Public Goods+Social Norm: “You belong to a minority because 94% of Belgian taxpayers filed on time. Why not follow their example? By filing and paying your taxes you guarantee the provision of essential services of the government, such as education, public health and public safety”

For this experiment, the main outcome is the probability of filing 21 days after letter receipt: 21 days is the time at which the tax administration uses its own estimate to calculate the tax liability.

Appendix Table A.8 presents the results of the experiment. We find no evidence that public goods messages, social norm messages, or the combination of the two increase the probability of filing on time. This is consistent with the results from the other experiments studied in this paper, which suggest that tax morale messages are not effective in raising tax compliance. However, in this case the complexity of the letter may have made the messages less salient.

A.2 Tax-on-web Survey

The answers to the following 10 questions are treated independently on an anonymous basis and are not linked to individual declarations.

1. On a scale of 1 to 10, to what extent do you find it easy to submit your tax return via Tax-on-Web?
2. On a scale of 1 to 10, how satisfied are you with the content and functions of Tax-On-Web?
3. On a scale of 1 to 10, how would you recommend Tax-On-Web to friend (s) or colleague (s)?
4. On a scale of 1 to 10, to what extent are you satisfied with the general tax system?
5. On a scale of 1 to 10, to what extent do you value the public services where (your) tax money is used for?
6. On a scale of 1 to 10, to what extent do you agree with the way your tax money is currently being spent?
7. On a scale of 1 to 10, to what extent do you think citizens should be completely honest when completing their tax return?
8. On a scale of 1 to 10, to what extent do you have a good idea of where your tax money goes?
9. Please add the following budget categories with the percentage of tax payable to you to these public services (total = 100%):
 - General government management (public debt, public services, basic research, foreign economic assistance, etc.)
 - Defence
 - Public order and safety
 - Economics
 - Environmental protection
 - Housing and common facilities
 - Recreation, culture and religion

- Education
- Health
- Social protection (elderly, sickness and disability, family and children, unemployment, ...)

10. If you had the opportunity to give your preference in terms of budget priorities, in which order would you spend the following categories on your tax money? Please place numbers from 1 (highest priority) to 10 (lowest priority) next to the following categories: (same as above)

Figure A.1: Tax morale treatment in filing process

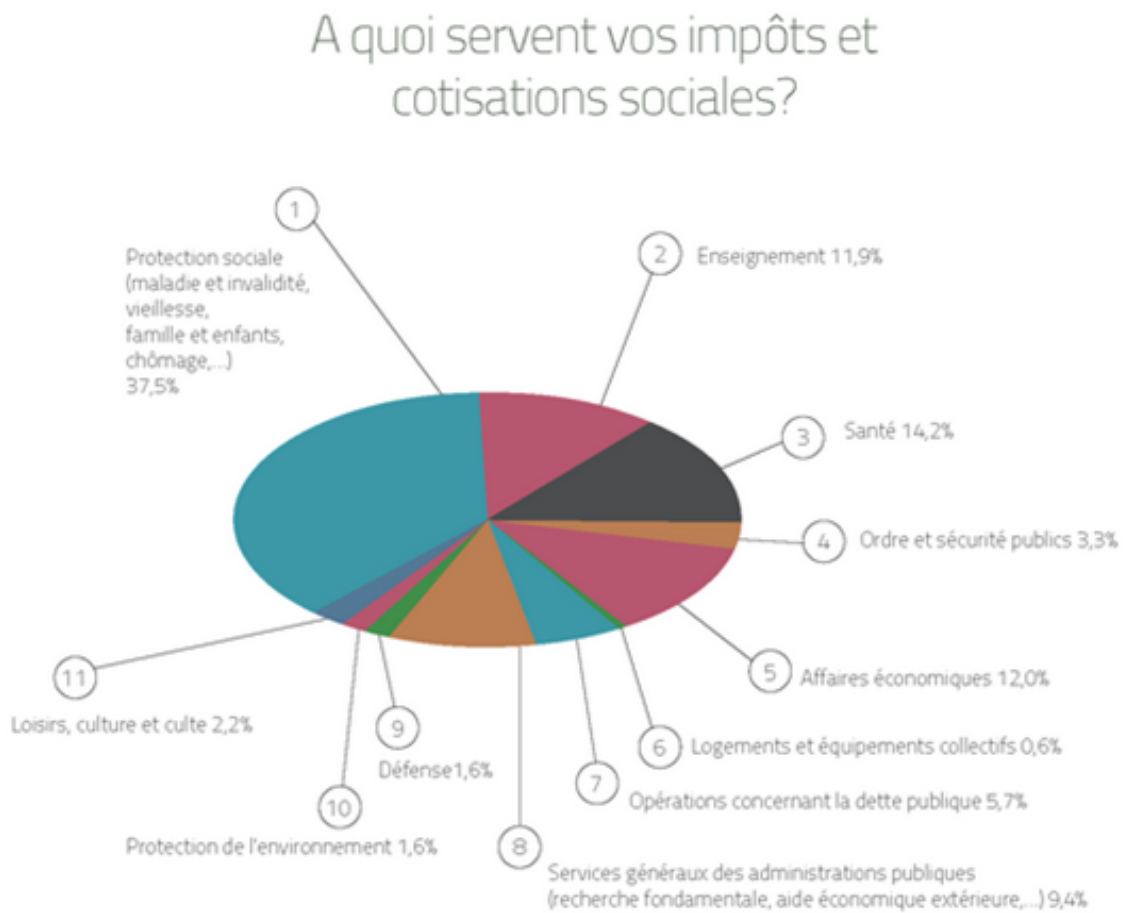
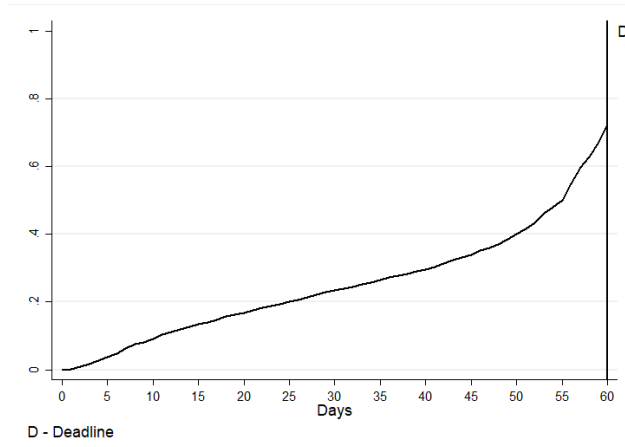
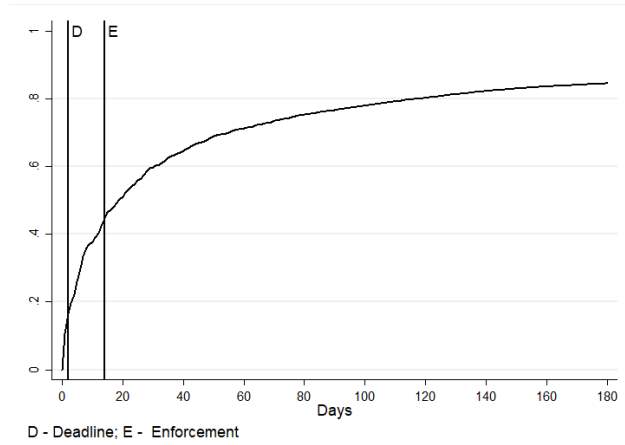


Figure A.2: Dynamics of Tax Compliance in the Control Group

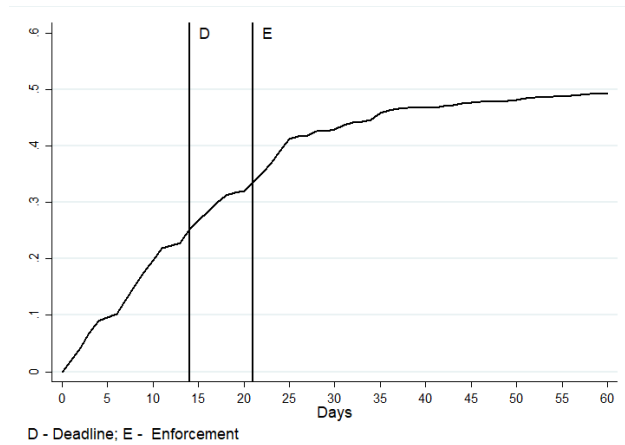
(a) Tax Payment



(b) Payment Reminder

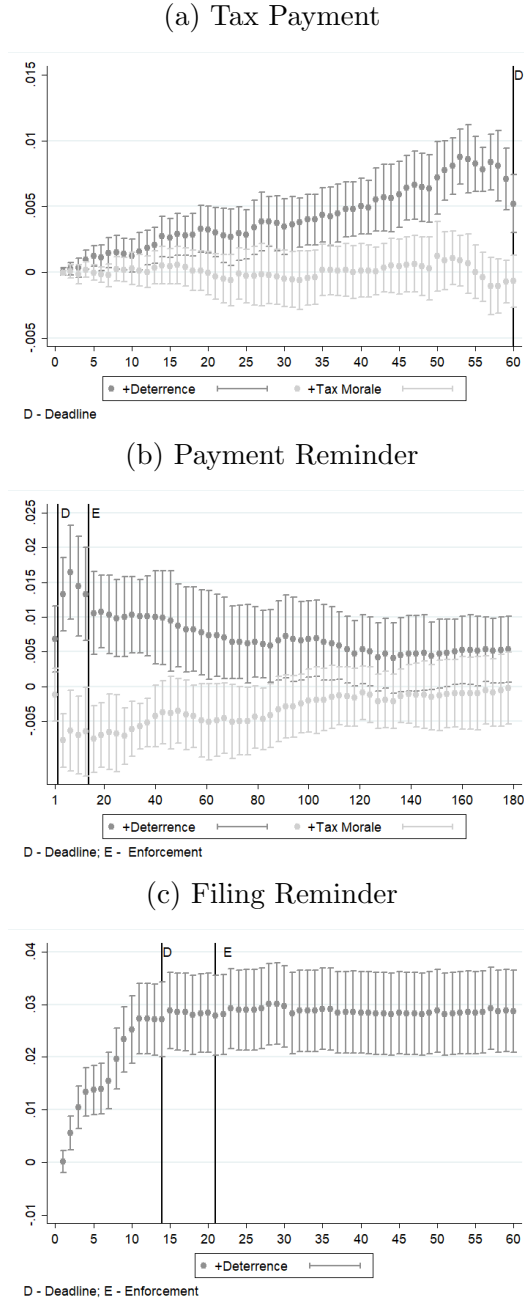


(c) Filing Reminder



Note: The figure presents average compliance in the control group by days since letter receipt for the TP (Panel (a)), TPR FY2014 (Panel (b)) and TFR FY2015 (Panel (c)) experiments. Outcome is partial payment probability at 60 days / deadline in Figure (a) and at 14 days / enforcement start in Figure (b); outcome is filing probability at 21 days / enforcement start in Figure (c).

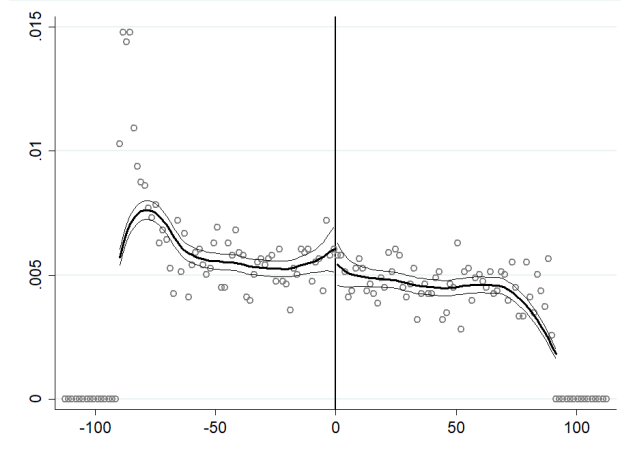
Figure A.3: Dynamic Effects of Deterrence and Tax Morale Messages



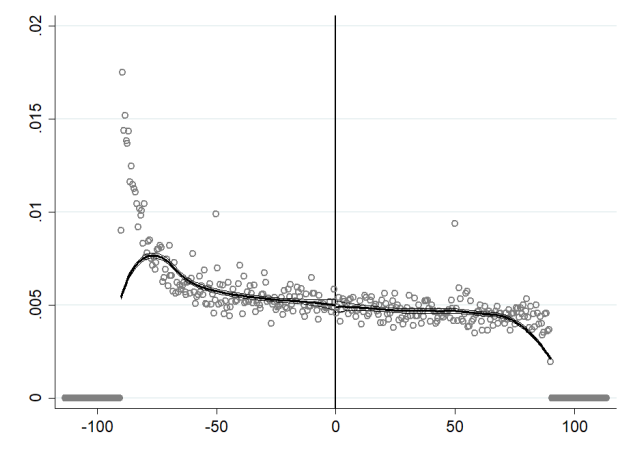
Note: The figure presents deterrence and tax morale treatment effect estimates by days since letter receipt for the TP (Panel (a)), TPR FY2014 (Panel (b)) and TFR FY2015 (Panel (c)) experiments. The outcome is partial payment probability in Panels (a) and (b), and filing probability in Panel (c). The vertical lines indicate the payment/filing deadline and/or the day follow-up enforcement starts. Controls are listed in Table 1. 95% confidence intervals based on robust standard errors are plotted. Standard errors are clustered by date of letter receipt in Panels (a) and (b).

Figure A.4: RDD – Identifying Assumptions

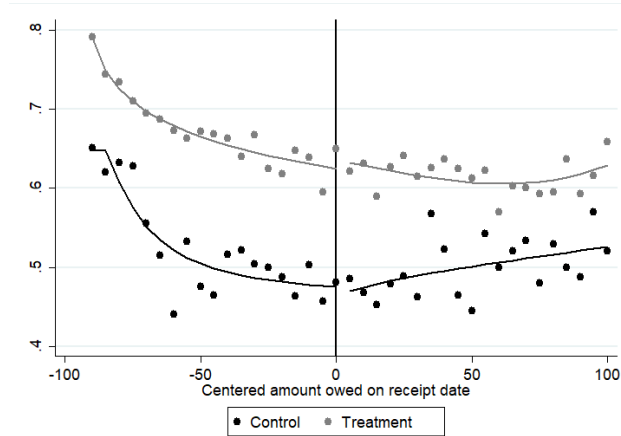
(a) Density around the threshold - Control



(b) Density around the threshold - Treatment



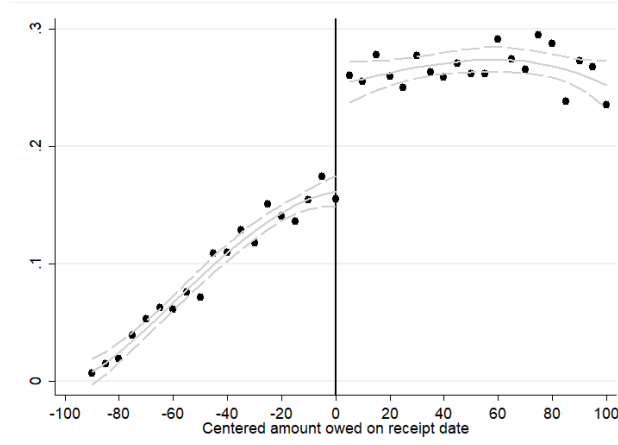
(c) Probability of Paying before Enforcement



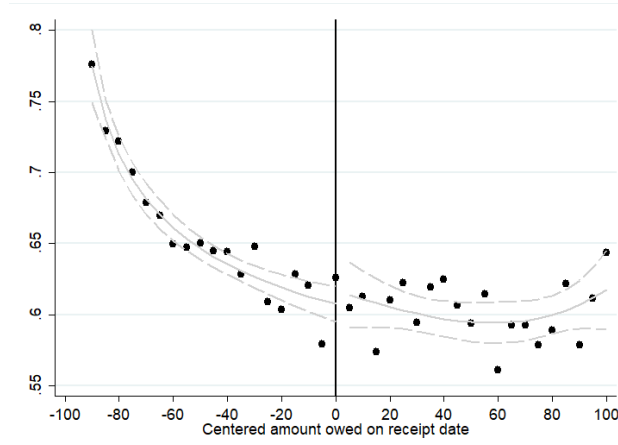
Note: The figure is based on the TPR FY2014 experiment. It explores the plausibility of the identification assumptions underlying the RDD. Panels (a) and (b) plot the average density by bin in the control and treatment group, respectively. Panel (c) plots the probability of payment before enforcement by initial amount owed (centred at the enforcement threshold). Bin size is set to €5 and amounts within €100 of the enforcement threshold are considered. Fractional polynomial predictions are plotted as well.

Figure A.5: Effects of Enforcement

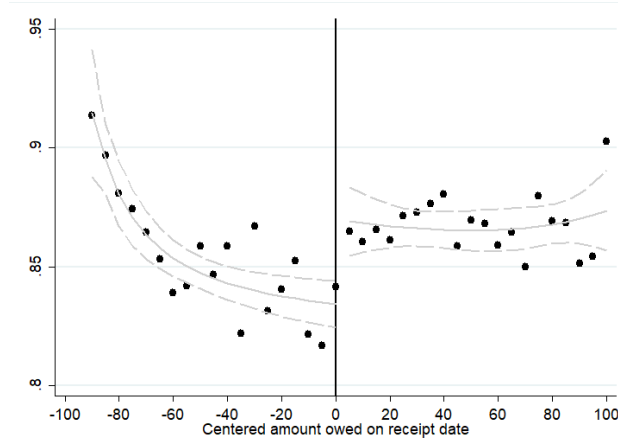
(a) Probability of Enforcement at 180 days



(b) Probability of Partial Payment at 14 days / enforcement start



(c) Probability of Partial Payment at 180 days



Note: The figure is based on the the TPR FY2014 experiment. It shows probability of enforcement after 180 days (Panel (a)), probability of paying after 14 days (Panel (b)) and probability of paying after 180 days (Panel (c)) by initial amount owed (centred at the enforcement threshold). Bin size is set to €5 and amounts within €100 of the enforcement threshold are considered. Fractional polynomial predictions with 95% confidence intervals plotted.

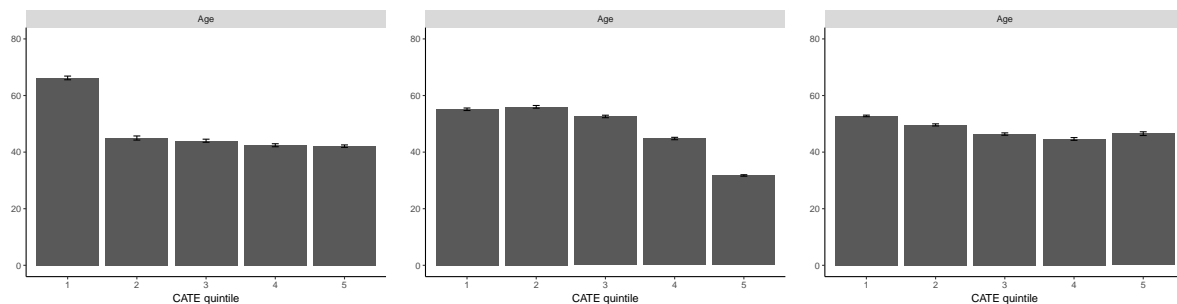
Figure A.6: Average Value of Control Variables by Quintile of Treatment Effects

Simplification

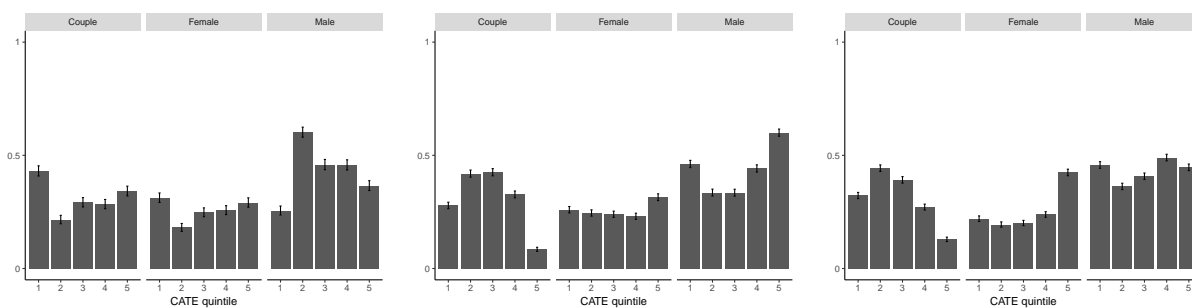
Deterrence

Tax Morale

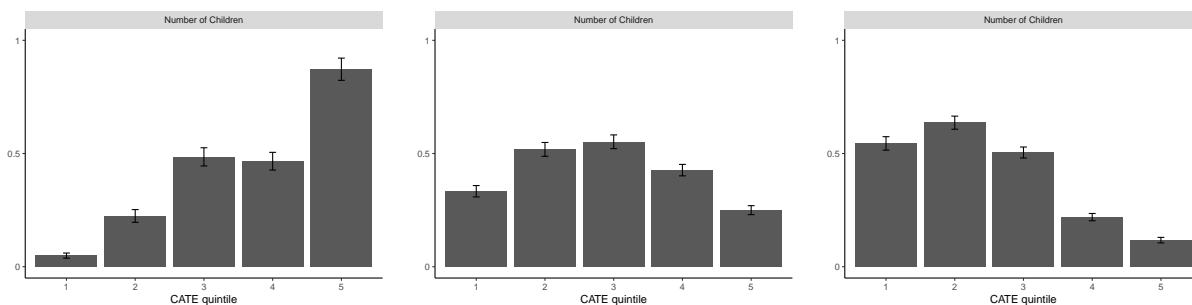
(a) Average Age



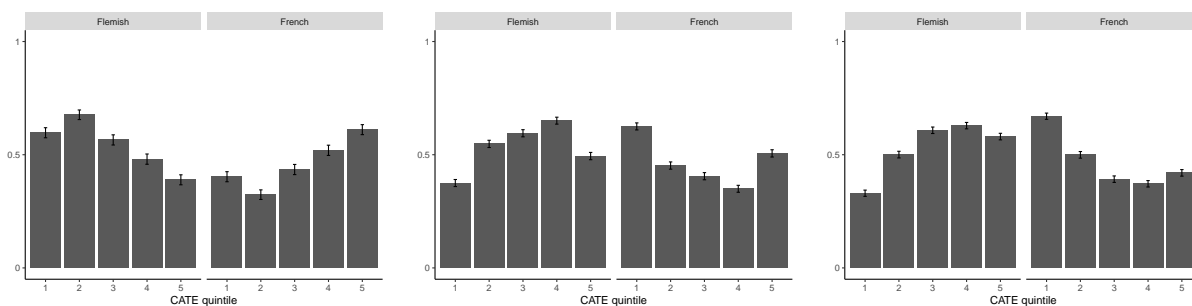
(b) Average of Gender categories



(c) Average Number of Children



(d) Average of Language Categories

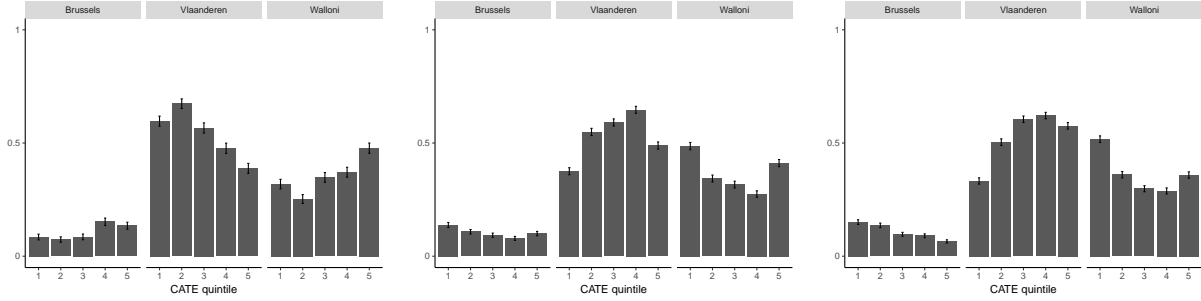


Simplification

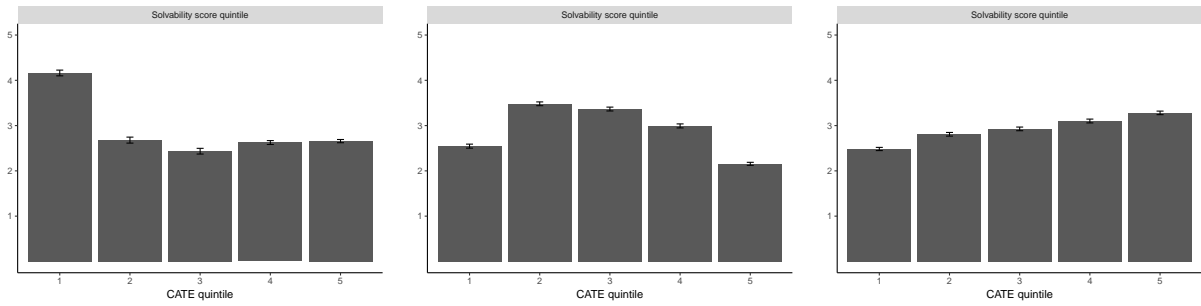
Deterrence

Tax Morale

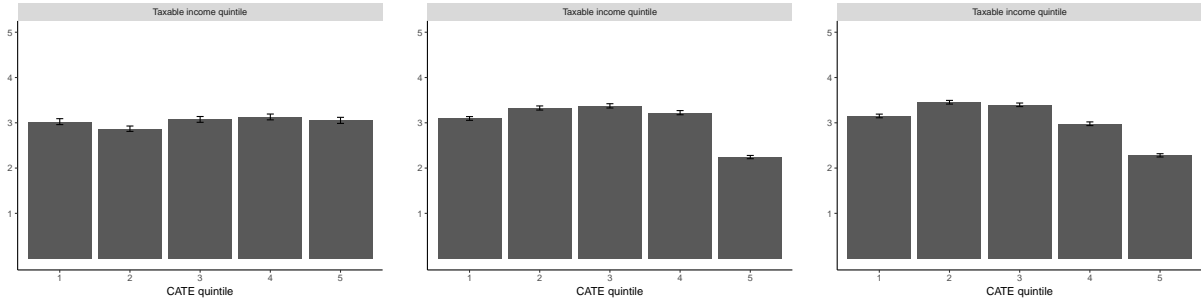
(e) Average of Region Categories



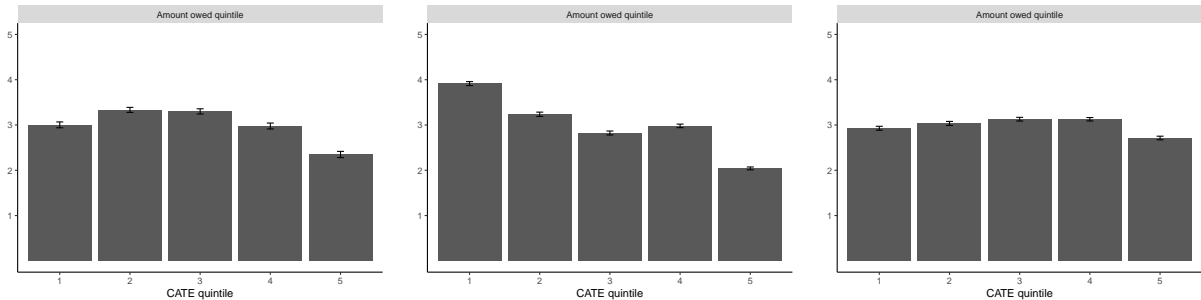
(f) Average Solvability Score



(g) Average Taxable Income



(h) Average Tax Liability



Note: The figure presents the mean and 95% confidence interval of control variables from TPR FY2014 experiment by quintile of conditional average treatment effect (CATE). These were estimated using the generalized random forest (GRF) algorithm ([Wager and Athey, 2018](#)). Three panels in each figure differ in the definition of treatment and control groups. The underlying sample of taxpayers are those in the control group and those sent a simplified letter without additional messages in the left panel, simplified letter and a simplified letter with a deterrence message in the middle panel, a simplified letter and a simplified letter with a tax morale message in right panel.

Table A.1: Letter Messages by Experiment

Experiment / Type	Name	Message
Panel A: Tax Payment		
Simplification	Not personalized	Madam, Sir (instead of taxpayer names)
Deterrence	Explicit Penalty	These costs amount to 209 euros on average and can go up depending on the circumstances.
	Enforcement + Immediacy	Warning: do not wait until the deadline to pay, you run the risk of being late. If you do not pay on time, we will start actions to recover this amount.
Tax Morale	Social Norm	In Belgium 95% of taxes are paid on time.
	Public Goods	Tax revenues allow basic public services such as health care, education and law and order, to function.
Panel B: Payment Reminders		
Deterrence	Explicit Penalty (FY2014, 2015)	These costs amount to 209.00 euro on average and may, depending on the situation, rise further.
	Active Choice (FY2014)	Not paying your taxes will be seen as an active choice.
	Explicit Penalty + Immediacy (FY2015)	These costs amount to 209.00 euro on average and may, depending on the situation, rise further. By paying now you may still avoid these costs.
Tax Morale	Explicit Penalty + Enforcement (FY2015)	These costs amount to 209.00 euro on average and may, depending on the situation, rise further. We will undertake actions to claim tax dues that may involve seizing your income or your assets.
	Explicit Penalty FM (FY2015)	Woman's name, Man's name (instead of reversed)
	Social Norm (FY2014, 2015)	You belong to a minority of taxpayers who did not pay their taxes within the legal period: 95% of taxes in Belgium are paid on time. Why not follow this example?
	Public Goods (FY2014)	Paying taxes guarantees the provision of essential services by the government, such as public health, education, and public safety.
	Public Goods Negative (FY2014, 2015)	Not paying taxes puts at risk the provision of essential services by the government, such as public health, education, and public safety.
Panel C: Tax Filing		
Tax Morale	Public Goods	The above pie chart illustrates how your taxes and social security contributions are spent in terms of public services.
	Public Goods Negative	The above pie chart illustrates how your taxes and social security contributions are spent in terms of public services. Incorrect and timely completion of the declaration puts at risk the essential services provided by the government.
	Public Goods + Penalty	The above pie chart illustrates how your taxes and social security contributions are spent in terms of public services. By completing your declaration correctly and in a timely fashion, you avoid further measures such as fines and tax increases.
	Public Goods + Social Norms	The above pie chart illustrates how your taxes and social security contributions are spent in terms of public services. The vast majority of people complete their declaration correctly and in a timely manner. Please follow this example.
Panel D: Filing Reminders		
Deterrence	Explicit Penalty (FY2015)	You risk a penalty of 50 to 1,250 euro and a tax increase of 10 to 200%.
Tax Morale	Social Norm (FY2014)	You belong to a minority as 94% of Belgians file their tax declarations on time. Why not follow this example?
	Public Goods (FY2014)	Paying taxes guarantees the provision of essential services by the government, such as public health, education, and public safety.

Note: The table lists all letter messages by experiment and treatment type.

Table A.2: Randomization Design

2-digits	TPR 2014	TFR 2015	TPR 2015	TF 2016	2-digits	TPR 2014	TFR 2015	TPR 2015	TF 2016	2-digits	TPR 2014	TFR 2015	TPR 2015	TF 2016
01	C	C	C	TM	51	TM	D	D	C	51	TM	D	TM	C
02	C	C	S	TM	52	TM	D	D	TM	52	TM	D	TM	C
03	C	C	D	TM	53	TM	D	D	TM	53	TM	D	TM	C
04	C	C	D	TM	54	TM	D	D	TM	54	TM	D	C	C
05	C	C	D	TM	55	TM	D	D	TM	55	TM	D	S	C
06	C	C	D	TM	56	TM	D	D	TM	56	TM	D	D	C
07	C	C	TM	TM	57	TM	D	D	TM	57	TM	D	D	C
08	C	C	TM	TM	58	TM	D	D	TM	58	TM	D	D	C
09	C	C	TM	TM	59	TM	D	D	TM	59	TM	D	D	C
10	C	C	TM	TM	60	TM	D	D	TM	60	TM	D	D	C
11	C	S	X	TM	61	TM	D	D	TM	61	TM	D	TM	C
12	S	S	C	TM	62	TM	D	D	TM	62	TM	D	TM	C
13	S	S	S	TM	63	TM	D	D	TM	63	TM	D	TM	C
14	S	S	D	TM	64	TM	D	D	TM	64	TM	D	X	C
15	S	S	D	TM	65	D	D	D	D	65	D	D	C	C
16	S	S	D	TM	66	D	D	D	D	66	D	D	S	C
17	S	S	D	TM	67	D	D	D	D	67	D	D	D	C
18	S	S	TM	TM	68	D	D	D	D	68	D	D	D	C
19	S	S	TM	TM	69	D	D	D	D	69	D	D	D	C
20	S	S	TM	TM	70	D	D	D	D	70	D	D	D	C
21	S	S	TM	TM	71	D	D	D	D	71	D	D	TM	C
22	S	D	X	TM	72	D	D	D	D	72	D	D	TM	C
23	TM	D	C	TM	73	D	D	D	D	73	D	D	TM	C
24	TM	D	S	TM	74	D	D	D	D	74	D	D	TM	C
25	TM	D	D	TM	75	D	D	D	D	75	D	D	X	C
26	TM	D	D	TM	76	D	D	D	D	76	D	D	C	C
27	TM	D	D	TM	77	D	D	D	D	77	D	D	S	C
28	TM	D	D	TM	78	D	D	D	D	78	D	D	D	C
29	TM	D	TM	TM	79	D	D	D	D	79	D	D	D	C
30	TM	D	TM	TM	80	D	D	D	D	80	D	D	D	C
31	TM	D	TM	TM	81	D	D	D	D	81	D	D	D	C
32	TM	D	TM	TM	82	D	D	D	D	82	D	D	D	C
33	TM	D	X	TM	83	D	D	D	D	83	D	D	TM	C
34	TM	D	C	TM	84	D	D	D	D	84	D	D	TM	C
35	TM	D	S	TM	85	D	D	D	D	85	D	D	TM	C
36	TM	D	D	TM	86	D	D	D	D	86	D	D	X	C
37	TM	D	D	TM	87	D	D	D	D	87	D	D	C	C
38	TM	D	D	TM	88	D	D	D	D	88	D	D	S	C
39	TM	D	D	TM	89	D	D	D	D	89	D	D	D	C
40	TM	D	TM	TM	90	D	D	D	D	90	D	D	D	C
41	TM	D	TM	TM	91	D	D	D	D	91	D	D	D	C
42	TM	D	TM	TM	92	D	D	D	D	92	D	D	D	C
43	TM	D	TM	TM	93	D	D	D	D	93	D	D	TM	C
44	TM	D	C	TM	94	D	D	D	D	94	D	D	TM	C
45	TM	D	S	TM	95	D	D	D	D	95	D	D	TM	C
46	TM	D	D	TM	96	D	D	D	D	96	D	D	TM	C
47	TM	D	D	TM	97	D	D	D	D	97	D	D	X	C
48	TM	D	D	TM		D								
49	TM	D	D	TM		D								
50	TM	D	TM	TM		D								

Note: The table presents jointly the randomization design of four separate experiments. TPR stands for payment reminders, TFR for filing reminders and TF for tax filing experiment. 2-digits stands for the last two digits of taxpayers' social security numbers. C stands for Control, S for Simplified letters, TX for Tax Morale messages and D for Deterrence messages.

Table A.3: Detailed Randomization Design - TPR (Payment Reminder) Experiments

[illegible]

Note: The table presents experimental design for payment reminders experiment FY2014 and payment reminders experiment FY2015. Numbers indicate last two digits of social security number of the taxpayer. Old - Old letter; New - Simplified letter; SN - Social Norm; PG-- Public Goods Neg; PPG - Public Goods Pos; SN PG - Social Norm Public Goods; AC - Active Choice; EP - Explicit Penalty; AC EP - Active Choice Explicit Penalty; EP FFM - Explicit Penalty Woman Man; EP+Enfo - Explicit Penalty+Enforcement; EP+Imm - Explicit Penalty+Immediacy; EP PG- - Explicit Penalty Public Goods Neg; EP SN - Explicit Penalty Social Norm

Table A.4: Overlap across experiments

Experiment	Share of taxpayers in experiment		
	Payment Reminders	Payment Reminders	Filing Reminders
	FY2014 (1)	FY2015 (2)	FY2015 (3)
Payment Reminders FY2014	1.000	0.283	0.062
Payment Reminders FY2015	0.307	1.000	0.066
Filing Reminders FY2015	0.106	0.104	1.000

Note: The table presents the overlap between populations of taxpayers in the payment reminders (TPR) and filing reminders experiments (TFR). Each cell gives the share of taxpayers in the experiment listed horizontally that were also part of the population of the experiment listed vertically.

Table A.5: Repeated Treatment Effects

	Probability of some payment 14 days (before enforcement) (1)
Simplified 2014	-0.00813 (0.0138)
Simplified 2015	0.0987*** (0.0117)
Simplified 2014 * Simplified 2015	0.00136 (0.0124)
+ Deterrence 2014	0.00817 (0.00557)
+ Deterrence 2015	0.00827 (0.00527)
+ Deterrence 2014 * + Deterrence 2015	-0.00365 (0.00720)
+ Tax Morale 2014	0.00214 (0.00729)
+ Tax Morale 2015	-0.0171*** (0.00272)
+ Tax Morale 2014 * + Tax Morale 2015	0.00145 (0.00699)
Control mean	0.423
N	64,736

Note: The table present treatment effect estimates for repeated treatment in the payment reminders experiment. Sample size is limited to individuals who were late with payment in both FY2014 and FY2015. For FY2015 treatment assignment both dummies for Deterrence and Tax Morale equal one for individuals who received a letter with both a deterrence and tax morale message. Control variables are listed in Table 1. Standard errors in parentheses are clustered by date of letter receipt. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.6: Payment Experiments: Individual Letter Effects

	Probability of some payment		
	14 days (before enforcement)		before the deadline
	TPR FY2014 (1)	TPR FY2015 (2)	TP (3)
<i>Simplification Treatments</i>			
Simplified	0.102*** (0.0103)	0.104*** (0.00479)	0.00464** (0.00162)
+ Not personalized			0.000841 (0.00125)
<i>Deterrence Treatments</i>			
+ Explicit Penalty	0.0204*** (0.00250)	0.0104** (0.00387)	0.00440*** (0.000893)
+ Active Choice	0.000774 (0.00364)		
+ Explicit Penalty+Active Choice	0.0160** (0.00549)		
+ Explicit Penalty+Enforcement		0.0252*** (0.00280)	
+ Explicit Penalty+Immediacy		0.0187*** (0.00407)	
+ Explicit Penalty FM		0.00971* (0.00479)	
+ Explicit Enforcement + Immediacy			0.00711*** (0.00124)
<i>Tax Morale Treatments</i>			
+ Public Goods Negative	-0.00741* (0.00366)	-0.0126*** (0.00262)	
+ Public Goods Positive	-0.0142*** (0.00378)		-0.00232 (0.00135)
+ Social Norms	-0.00232 (0.00377)	-0.01000** (0.00404)	0.000841 (0.00125)
+ Social Norms+Public Goods Positive	-0.00646 (0.00432)		
<i>Deterrence & Tax Morale Treatments</i>			
+ Explicit Penalty+Social Norm		0.00769** (0.00311)	
+ Explicit Penalty+Public Goods Negative		0.00601 (0.00448)	
Control mean	0.447	0.419	0.728
N	229,751	188,180	1,216,317

Note: The table presents treatment effect estimates of messages in the two payment reminder experiments (TPR 2014 in column 1 and TPR 2015 in column 2) and in the tax payment (TP) experiment (column 3). Control variables are listed in Table 1. Standard errors in parentheses are clustered by date of letter receipt.

* p<0.1; ** p<0.05; *** p<0.01.

Table A.7: Treatment Effects on Other Outcomes

Panel A: Tax Payment	TPR 2014	TPR 2015	TP
	% Liability Paid before Enforcement (1)	% Liability Paid before Enforcement (2)	% Liability Paid before Deadline (3)
Simplified	0.00307 (0.00183)	0.00995** (0.00404)	0.00162** (0.000599)
+ Deterrence	0.00210 (0.00149)	0.00227 (0.00185)	-0.000238 (0.000507)
+ Tax Morale	0.000384 (0.00156)	-0.000729 (0.00224)	-0.000447 (0.000547)
+ Deterrence & Tax Morale		0.00327 (0.00218)	
Control mean	0.915	0.900	0.941
N	124,032	98,422	892,310
Panel B: Tax Filing	Log pre-check total tax due (1)	Log self-employed profits (2)	Log self-employed expenses (3)
Tax Morale	-0.00265 (0.00257)	0.000692 (0.00789)	0.000189 (0.00774)
Control mean	13.45	12.77	12.94
N	850,778	51,183	36,260
Panel B (continued)	Log salaried expenses (4)	Log general expenses (5)	
Tax Morale	-0.00404 (0.00335)	-0.00526 (0.00372)	
Control mean	13.15	11.08	
N	33,103	233,889	

Note: The table presents treatment effect estimates for other outcomes of interest in the tax payment (TP FY2016 Panel A) and the tax filing (TF FY2016 Panel B) experiments. In Panel A the sample consists of late payers who had made some payment before enforcement started. Control variables are listed in Table 1. Robust standard errors in parentheses, clustered by date of letter receipt in Panel A. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.8: Filing Reminders FY2014

	Probability of having filed 21 days (before enforcement) (1)
Public Goods	-0.00502** (0.00243)
Social Norms	0.000550 (0.00244)
PG+SN	-0.00263 (0.00244)
Control mean	0.147
N	162,682

Note: The table presents treatment effect estimates from the filing reminders experiment (TFR FY2014). Control variables are listed in Table 1. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.9: Payment Reminders FY2014

	Probability of some payment			
	2 days (1)	14 days (2)	30 days (3)	180 days (4)
Simplified	0.0645*** (0.0114)	0.103*** (0.0102)	0.0688*** (0.00425)	0.00954** (0.00345)
Control mean	0.166	0.447	0.598	0.845
N	229,751	229,751	229,751	229,751

Note: The table presents treatment effect estimates from the payment reminders experiment (TPR FY2014). Control variables are listed in Table 1. Standard errors in parentheses are clustered by date of letter receipt. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.10: Number of Follow-up Enforcements FY2014

	Nr registered letters within 180 days (1)	Nr garnishments within 180 days (2)	Nr bailiffs within 180 days (3)
Simplified	-0.0740*** (0.00272)	-0.0278*** (0.00206)	-0.0118*** (0.00157)
Control mean	0.350	0.134	0.078
N	229,751	229,751	229,751

Note: The table presents treatment effect estimates on the number of enforcement actions by type from the payment reminders experiment (TPR FY2014). Control variables are listed in Table 1. Standard errors in parentheses are clustered by date of letter receipt. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.11: RDD: Number of Follow-up Enforcements FY2014

	Nr registered letters within 180 days (1)	Nr garnishments within 180 days (2)	Nr bailiffs within 180 days (3)
Simplified	-0.0670*** (0.0173)	-0.0235* (0.0135)	0.00223 (0.00329)
Enforcement	0.102*** (0.0236)	0.0592*** (0.0184)	-0.000523 (0.00447)
Simplified*Enforcement	-0.0506** (0.0252)	-0.0162 (0.0197)	0.000178 (0.00477)
Control mean	0.154	0.056	0.002
N	28,665	25,246	29,409

Note: The table presents treatment effect estimates from the regression discontinuity design analysis embedded in the payment reminder experiment (TPR FY2014). Simplified is a dummy variable equal to one for taxpayers who received a simplified letter. Enforcement is a dummy variable equal to one for liability amounts above the cut-off value. Control variables are listed in Table 1. Standard errors in parentheses are clustered by date of letter receipt. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.12: Heterogeneous Effects – Payment Reminders Experiment FY2015

Probability of payment at 14 days (before enforcement)			
Variable	(1)	Variable	(2)
Simplified	0.0939*** (0.0274)	Solvency score Q2 * Simplified	0.0502* (0.0265)
+ Deterrence	0.0220 (0.0232)	* Deterrence	0.00372 (0.0178)
+ Social	-0.0283 (0.0207)	* Social	0.00544 (0.0151)
Age 31-40 * Simplified	0.00507 (0.0132)	Solvency score Q3 * Simplified	0.0862*** (0.0263)
* Deterrence	-0.00330 (0.0147)	* Deterrence	-0.0285* (0.0155)
* Social	-0.0109 (0.0113)	* Social	-0.0302* (0.0150)
Age 41-50 * Simplified	-0.0188* (0.0106)	Solvency score Q4 * Simplified	0.0399* (0.0211)
* Deterrence	-0.00424 (0.00923)	* Deterrence	-0.00731 (0.0127)
* Social	-0.00785 (0.0139)	* Social	-0.00416 (0.0153)
Age 51-60 * Simplified	-0.0192 (0.0180)	Solvency score Q5 * Simplified	0.0282 (0.0168)
* Deterrence	0.000563 (0.0146)	* Deterrence	-0.00214 (0.0143)
* Social	-0.00704 (0.0166)	* Social	0.0126 (0.0134)
Age 61+ * Simplified	-0.0326** (0.0134)	Liability Q2 * Simplified	-0.0327** (0.0140)
* Deterrence	-0.00480 (0.0144)	* Deterrence	-0.00550 (0.0116)
* Social	-0.00792 (0.0127)	* Social	0.0256* (0.0129)
One child * Simplified	0.0143 (0.0118)	Liability Q3 * Simplified	-0.0158 (0.0163)
* Deterrence	0.00540 (0.0129)	* Deterrence	-0.00686 (0.00745)
* Social	0.0195 (0.0158)	* Social	0.0102 (0.00675)
Two or more children * Simplified	0.0379** (0.0168)	Liability Q4 * Simplified	-0.0518** (0.0218)
* Deterrence	-0.0136 (0.0124)	* Deterrence	-0.0136 (0.00986)
* Social	-0.0169 (0.0103)	* Social	0.0202** (0.00829)
		Liability Q5 * Simplified	-0.0540** (0.0200)
		* Deterrence	-0.0223** (0.00931)
		* Social	0.0205 (0.0119)
Control mean			0.333
N			188,180


Note: The table presents treatment effect estimates from the heterogeneous effects analysis of the second payment reminder experiment (TPR FY2015). Control variables are listed in Table 1. The full set of interactions between individual control and treatment variables are included in the estimation (coefficients not reported). Estimates for Deterrence and Tax Morale joint treatment omitted for brevity. Standard errors in parentheses are clustered by date of letter receipt. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.13: Filing Reminders FY2015

	Probability of having filed 21 days (before enforcement) (1)
Simplified	0.0191* (0.0105)
+ Deterrence	0.0284*** (0.0102)
Control mean	0.317
N	148,925

Note: The table presents treatment effect estimates from filing reminders experiment (TFR FY2015). Control variables are listed in Table 1. Additional controls include dummies for the treatment the taxpayer would have received if had been late with payment in the previous fiscal year. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Letter A.9: Payment Reminder Experiment - Old Letter



Service Public

Fédéral

FINANCES

Exp.: TEAM REC PP NIVELLES 2

BD DES ARCHERS 71 1400 NIVELLES

RECOUVREMENT ET PERCEPTION

Tél : 0257/595.00

IBAN : BE58679200235379

BIC : PCHQBE33

NOS REFERENCES : Article

Commune : TUBIZE

IL A ETE TENU COMPTE DES PAIEMENTS EFFECTUES JUSQU'AU : 8.11.2016

IMPOT DES PERSONNES PHYSIQUES EXERCICE D'IMPOSITION 2016

Madame, Monsieur,

Sauf erreur de ma part, la cotisation citée en référence est restée impayée à concurrence de :

Impôt dû : 255,85

Frais dus : 0,00

Intérêts : 0,00

Montant déjà payé (impôt + frais + intérêts) : 0,00

Solde restant dû à ce jour : 255,85 EUR

Je vous invite à régler immédiatement le solde restant dû. A défaut de paiement immédiat au moyen de la formule ci-dessous, je me verrais malheureusement contraint de poursuivre le paiement de ce solde par tous les moyens que la loi met à ma disposition, ce qui entraînerait des frais supplémentaires qui seront mis à votre charge.

TAUX D'INTERET=9,6% JUSQU'AU 31/12/98, ENSUITE 7%

Vous trouverez au verso divers renseignements utiles.

Veillez agréer, Madame, Monsieur, l'assurance de ma considération distinguée.

Le Receveur

RENSEIGNEMENTS UTILES

1. AU CAS OU VOUS AURIEZ DEJA EFFECTUE LE PAIEMENT RECLAME, veuillez sans délai m'en informer et me faire parvenir la preuve du paiement (original ou photocopie).

2. N'ATTENDEZ PAS LE DERNIER JOUR DU DELAI pour effectuer votre paiement : un ordre de paiement ne produit ses effets que quelques jours ouvrables après sa remise à l'organisme financier.

3. SI VOUS ESTIMEZ DEVOIR MODIFIER LE MONTANT OU SI VOUS SOUHAITEZ NE PAS UTILISER LA FORMULE DE PAIEMENT ATTACHEE A LA SOMMATION vous devez reproduire scrupuleusement la COMMUNICATION qui figure sur cette formule de paiement.

4. MOYENS DE RECOUVREMENT PREVUS PAR LA LOI.
A défaut de paiement spontané de l'impôt, le receveur peut notamment saisir vos rémunérations et vos biens ; le cas échéant, il peut faire vendre ces derniers.

5. En cas de contestation de cette imposition, vous n'êtes tenu au paiement immédiat que du montant qui vous a été communiqué :
- soit par le fonctionnaire de taxation chargé de la régularisation de l'erreur constatée ;
- soit par le fonctionnaire chargé de l'instruction de votre réclamation (avis 178 J). Toutefois, dans ce cas, le receveur peut en outre prendre toutes les mesures conservatoires (hypothèque, saisie conservatoire...) destinées à garantir le paiement ultérieur du solde (frais et intérêts compris).

ORDRE DE VIREMENT

03

Si complété à la main, n'indiquer qu'une seule MAJUSCULE ou un seul chiffre noir (ou bleu) par case

Montant

EUR CENT

255,85

Compte donneur d'ordre (IBAN)

Compte bénéficiaire (IBAN)

BIC bénéficiaire

Nom et adresse bénéficiaire

Communication

Nom et adresse donneur d'ordre

TEAM REC PP NIVELLES 2

BD DES ARCHERS 71

1400 NIVELLES

67

Letter A.10: Payment Reminder Experiment - Simplified Letter

16.11.2016



**Service Public
Fédéral
FINANCES**

Exp.: TEAM REC PP NIVELLES 2
BD.DES ARCHERS 71 1400 NIVELLES

Administration générale de la Perception et du Recouvrement

Pour plus d'information, vous pouvez vous adresser à :
Tél. : 0257 595 00
E-mail : teamrec.nvelles2@minfin.fed.be
Heures de bureau de 9 h à 12 h ou sur rendez-vous

Votre numéro de dossier :

Madame xxx xxx,

Vous avez une dette fiscale impayée (voir verso).

Veillez payer **274,50 euros** endéans les **48 heures** sur le compte **BE58 6792 0023 5379** avec la communication structurée **++++000/0000/00000000++**.

Si vous ne payez pas, nous engagerons les procédures pour récupérer ce montant. Les intérêts de retard et les frais sont à votre charge.

Si, entre-temps, vous avez effectué le paiement, nous vous en remercions.

Salutations distinguées,

Le conseiller recouvrement - receveur

Num	Numéro national	IMPOT DES PERSONNES PHYSIQUES	
Xxx xxx	Nature de l'impôt	2016	
	Exercice d'imposition		
	Article de rôle		
<hr/>			
Décompte			
	Impôt établi	274.50	
	Frais	0,00	
	Intérêts	0,00	
	Montant déjà payé* (impôt + frais + intérêts)	- 0,00	
	Solde restant dû à ce jour		€ 274,50
<hr/>			
* IL A ETE TENU COMPTE DES PAIEMENTS EFFECTUES JUSQU'AU : 8/11/2016			

Les intérêts de retard sont dus par mois, au taux annuel de 7%.

Pour plus d'information : www.finances.belgium.be avec le mot clé 'payer' dans le champ de recherche.

Si vous avez introduit une réclamation, contactez-nous par e-mail (teamrec.niveles2@minfin.fed.be) ou par téléphone (0257 595 00). Pour plus d'information : www.finances.belgium.be avec le mot clé 'réclamation' dans le champ de recherche.

[illegible]

Letter A.11: Filing Reminder Experiment - Old Letter



Service Public
Fédéral
FINANCES

Administration générale de la
Fiscalité

Exp. : Mentionnez l'adresse de votre service

vos références	nos références	annexe(s)
----------------	----------------	-----------

Rappel : Avez-vous oublié de rentrer votre déclaration à l'impôt des personnes physiques pour l'exercice d'imposition 2016 (revenus 2015) ?

Madame, Monsieur,

Au 01.09.2016, nous n'avons **pas encore reçu votre déclaration** à l'impôt des personnes physiques pour l'exercice d'imposition 2016 (revenus 2015). La date limite de rentrée dans les délais était le 30.06.2016.

Le non-dépôt ou le dépôt tardif de la déclaration constitue une infraction. Les sanctions qui peuvent vous être appliquées sont reprises au verso de cette lettre.

Vous pouvez éviter ces sanctions en rentrant encore **vosre déclaration dans un délai de 14 jours** à partir de l'envoi de cette lettre. Vous trouverez la procédure pour le faire au verso.

Votre déclaration reste néanmoins tardive. Ce rappel ne modifie en rien le délai de rentrée légal initial. Si des motifs ou circonstances graves vous ont empêché de rentrer la déclaration au plus tard le 30.06.2016, vous devez les communiquer par écrit à votre bureau de taxation.

Vous ne devez pas réagir à cette lettre si :

- vous avez entretenu votre déclaration ;
- vous avez obtenu un délai supplémentaire valable pour rentrer votre déclaration après le 01.09.2016 ;
- vous passez par un mandataire pour rentrer votre déclaration. Votre mandataire peut encore rentrer votre déclaration jusqu'au 29.10.2016 inclus.

Veuillez agréer, Madame, Monsieur, nos salutations distinguées,

Le chef de service

Comment pouvez-vous encore rentrer votre déclaration ?

Deux possibilités :

- via www.taxonweb.be.
A cet effet, vous avez besoin de votre carte d'identité électronique et d'un lecteur de carte ou d'un token (pour chaque partenaire dans le cas d'une déclaration commune).

- en envoyant le formulaire de déclaration papier au :

SPF Finances - Centre de scanning
BP 51000
5100 Jambes

N'oubliez pas de dater et de signer ce formulaire de déclaration (par les deux partenaires dans le cas d'une déclaration commune).

Si vous n'avez pas reçu votre formulaire de déclaration ou si vous l'avez perdu, vous pouvez demander un exemplaire auprès de votre bureau de taxation :

AdminName1 – Phone – Email

Quelles sanctions risquez-vous ?

Si vous ne rentrez pas ou tardivement votre déclaration à l'impôt des personnes physiques, le SPF Finances peut :

- appliquer des **sanctions administratives** comme :
 - une amende administrative de 50 à 1250 euros (article 445, CIR 92) ;
 - un accroissement d'impôt de 10 % à 200 % (article 444, CIR 92) ;
- établir l'impôt durant un **délai d'imposition de 3 ans** (article 354, alinéa premier, CIR 92) ;
- appliquer la procédure de **taxation d'office** (article 351, CIR 92) ;
- pour les entreprises et les titulaires de profession libérale, appliquer le « **montant minimum des bénéfices ou profits imposables** » (article 342, § 3, CIR 92).

Avez-vous encore des questions ?

Pour plus d'informations sur votre dossier, vous pouvez prendre contact avec votre bureau de taxation :

AdminName1 – Phone – Email

Letter A.12: Filing Reminder Experiment - Simplified Letter



Service Public
Fédéral
FINANCES

Administration générale de la
Fiscalité

Exp. : Mentionnez l'adresse de votre service

M. JAN PEETERS
Mme. PETRA JANSENS
KERKSTRAAT 1
1000 BRUSSEL

vos références	vos références	annexe(s)
----------------	----------------	-----------

Madame Jansens,
Monsieur Peeters,

Nous n'avons pas encore reçu votre déclaration à l'impôt des personnes physiques pour l'exercice d'imposition 2016 (revenus 2015). Cependant, vous deviez la renvoyer pour le 30.06.2016.

Veillez renvoyer votre déclaration dans un délai de 14 jours.

Si non, vous risquez une amende de 50 à 1.250 euro et un accroissement d'impôt de 10 à 200%.

Comment renvoyer votre déclaration?

- via taxonweb.be, ou
- en envoyant le formulaire de déclaration papier au:
SPF Finances - Centre de scanning
BP 51000
5100 Jambes

N'oubliez pas de dater et de signer ce formulaire de déclaration (par les deux partenaires dans le cas d'une déclaration commune).

Vous ne devez pas réagir à cette lettre si :

- vous avez déjà déclaré votre déclaration ;
- vous avez obtenu un délai supplémentaire ;
- vous passez par un mandataire pour renvoyer votre déclaration.

Encore des questions?

AdminName1 – Phone – Email

Cordialement,

Le chef de service