What motivates people to pay their taxes? Evidence from four experiments on tax compliance*

Eric Floyd
University of California San Diego

Michael Hallsworth
Imperial College London
The Behavioural Insights Team

John A. List
University of Chicago and NBER

Robert D. Metcalfe
University of Southern California and NBER

Kristian Rotaru
Monash University

Ivo Vlaev
University of Warwick

Abstract

In this study, we first present a large natural field experiment that tested messages aimed at increasing tax compliance. We find that the main drivers of changes in compliance are messages describing the monitoring and enforcement behavior of the tax collector. A second natural field experiment built on the results of the first experiment to further investigate what kinds of costs resulting from tax collector oversight are salient to taxpayers. Specific time and cognitive incentives did not significantly increase payment rates, whereas stating non-specific costs of inaction did. Additional analyses suggest the increase in compliance is likely due to a 'fill in the blank' effect in which taxpayers assume the consequence is a fine. Interestingly, specifically stating maximum fine or jailtime consequences have the largest effect in a laboratory setting but only if the consequences are interpreted as realistic. Overall, our study reinforces that tax authorities can use short messages to increase tax compliance; the estimated accelerated revenue from the two field studies amounts to £9.9m.

Keywords: taxation; compliance; behavioral economics; natural field experiment; accounting.

JEL: C93; D03; H26

* We would like to thank Her Majesty's Revenue and Customs and the UK Cabinet Office's Behavioural Insights Team. In particular, Nick Down, Pedro Wrobel, David Halpern, Jackie Simms, Lesley Titley, Matt Cole, Dawn Reade, Simon Bradford, Mike Holgate, Bob Appleton, Gordon Smith, Graham Brammer, Mike Drewery, and Ara Darzi. We would also like to thank seminar participants in the accounting groups at the University of Minnesota and CUHK-Shenzhen as well as participants of the 3rd Monash Behavioural Business Laboratory Symposium. We also thank the Monash Business School and the Australian Research Council (ARC LIEF grant LE130100112) for funds associated with the Monash Business Behavioural Laboratory equipment used in this study. Floyd: ejfloyd@ucsd.edu; Hallsworth: m.hallsworth11@imperial.ac.uk; List: jlist@uchicago.edu; Metcalfe: robert.metcalfe@usc.edu; Rotaru: kristian.rotaru@monash.edu; Vlaev: ivo.vlaev@wbs.ac.uk.

1. INTRODUCTION

Understanding how to motivate individuals and organizations to pay their taxes has become a major issue for academic research and public policy (Andreoni et al., 1998). The volume and variety of tax compliance research has expanded greatly over recent decades (Sandmo, 2005). The earliest studies of Allingham and Sandmo (1972) and Srinivasan (1973) proposed a simple yet elegant expected utility model, with the decision to evade tax being based on income, tax rates, size of fine, and probability of audit. Many studies over the past thirty years have attempted to demonstrate and quantify the effect of these parameters (Andreoni et al., 1998). The result has been a much more sophisticated view of the economic determinants of tax compliance, which is of increasing interest to policy makers (Aaron and Slemrod, 2004, OECD, 2010).

Allingham and Sandmo (1972) did, however, also note that "other factors" (such as reputational concerns) were likely to affect the compliance decision. Since then, an increasing number of studies have attempted to identify, measure, and analyze the effect of factors such as social norms, fairness and moral concerns (Alm, 2012). A notable aspect of these studies is that they explicitly draw on other disciplines, particularly psychology, to explain taxpayer behavior (Kirchler, 2007). They also suggest varying routes by which these other factors affect behavior, with some incorporating them into a utility function, and others rejecting the idea that they can be weighted and ranked in this way (Gordon, 1989; Kirchler, 2007). Overall, academics have theoretically identified many potential factors that affect tax compliance, but little empirical consensus has resulted on which matter in reality.

In this study, we build on the existing literature in contemporary tax compliance research, (Alm & Martinez-Vazquez, 2003; McGraw & Scholz, 1991; Smith, 1992) by constructing messages intended to increase tax compliance. Our natural field experiments consisted of varying the wording of 300,000 tax payment reminders sent by the UK tax authority to those who had failed to pay income tax on time and measuring the effect these changes have on payment rates. The first experiment, conducted in 2012, tested multiple treatments. One group of messages made salient the degree of oversight being exerted by the tax authority; a second group included messages representing moral concerns, public goods, and a supportive approach. Many of the variables used in the second group have been included in prior studies; their inclusion allowed us to compare our study to prior work. Finally, two messages combined elements from both sets of approaches. None of the treatments from the second group significantly increased the rate of payments made within

three weeks of the letter issue.[1] In contrast, all the messages from the first group significantly increased payments, the most successful by five percentage points. These findings counter the view that inexpensive mailed messages have little or no impact on tax compliance (Blumenthal et al., 2001).

The second experiment was conducted twelve months later, in 2013. It aimed to build on the findings of the first experiment by considering oversight costs in more depth. The messages concern economic costs (interest charges), time, and cognitive costs, and the cost of potential further action by the authorities. We find no significant effect on payment rates from messages based on economic, time and cognitive costs. Messages regarding further enforcement action significantly increased payment rates (again, by up to five percentage points), although the effects appeared to vary according to message framing.

Our third and fourth experiments were conducted using Prolific and, in the lab, respectively. The goal of the experiments was to further understand why the message regarding further enforcement action was effective in experiment two. We find that the further enforcement action message in experiment two is about as effective as messages stating non-specific or average fines. This suggests that taxpayers in experiment two may be "filling in the blanks" and assuming the further enforcement action is a fine. However, we also find that if the message mentions a severe fine or jailtime payment rates can increase even further.

Our paper contributes to the tax compliance literature in several ways. First, in Hallsworth et al. (2017), none of the treatments focused on the expected costs of getting caught. Our current paper has a monitoring treatment that explicitly states that the taxpayer is going to be closely monitored and thus raises the expected cost of not paying taxes. We can compare this cost treatment with that from the moral cost treatment groups in the same experiment, which has not been done before in the tax compliance literature. Interestingly, we find that moral cost messages are sensitive to precise wording and framing. Therefore, monitoring messages may be a more reliable way to encourage tax payment.

Our finding that oversight from the tax authority changes compliance behavior is consistent with prior literature, which alter the audit probability of the tax collector. However, there has been less focus on what costs taxpayers are worried about and whether the framing of these costs matter. That is, most studies focus on the probability of monitoring as opposed to the potential consequences of monitoring. Our focus in this paper is on the potential consequences. Our primary conclusion is

[1] It must be noted that we did not test descriptive or injunctive social norms alone in this paper, but they have been shown to be important in previous work (see Hallsworth et al., 2017).

that costs of further action from the tax authority can be a strong motivator for paying taxes. Time and cognitive costs, on the other hand, do not appear to be as effective.

Our paper also suggests that vaguely defined consequences can be effective at motivating tax compliance as specifically defined consequences. The mechanism behind this appears to be the fact that vaguely defined consequences leave the possibility open that taxpayers "fill in the blank" with the most likely consequences. In our setting, taxpayers appear to assume that vaguely defined consequences imply a fine. Policy makers may consider this type of wording when committing to certain punishments is not a possible policy lever. We further show that the tax authority could potentially increase compliance even more if they were able to threaten jailtime. Thus, it is also possible that vague messages are effective because they leave open the possibility of worst-case scenarios.

This study has policy implications: it reinforces that tax authorities can use short messages to increase tax compliance; the estimated accelerated revenue from the two studies amounts to £9.9m. It should be noted that these benefits were achieved with very small costs (amounting to the opportunity costs incurred by tax officials), since the reminder letters would have been issued regardless. We recommend that policy makers explore the extent to which similar inexpensive mailings could be used to increase tax compliance, perhaps also in settings encouraging small business tax compliance.

2. LITERATURE REVIEW

As Alm (2012) and Floyd and List (2016) note, economic studies are increasingly turning to experimental methods, partly because of the measurement difficulties associated with more traditional empirical analyses (Bloomfield, Nelson, and Soltes, 2016). Accounting research is following this trend as well. The great majority of experimental methods within tax compliance have been applied in laboratory settings: participants generally declare "income" over repeated rounds, under the threat of a fine for non-compliance, and receive their net gain at the end of the process (Webley, 1991). Taking an experimental approach offers major advantages. For example, it allows researchers to estimate the effect of crucial aspects (like the tax rate or penalty rate) that are very difficult to vary exogenously in the real world, and it provides precise measurements of non-compliance (Alm & Jacobson 2007).

Despite these advantages there are questions about the extent to which these laboratory findings translate into the real world (Elffers et al., 1992), which mirror debates in the field of economics more generally (Harrison & List, 2004; Levitt and List, 2007). This is of concern to

accountants, who are particularly concerned with how economic insights apply in "real world" institutions. Economists and accountants have increasingly been turning to field experiments to provide such "real world" evidence (List, 2009; Floyd & List, 2016). However, for a variety of reasons, there have been relatively few field experiments concerning tax compliance (Slemrod & Weber, 2012).[2]

Turning away from methodological issues concerning past research, tax compliance has been of interest to several fields of inquiry. From the perspective of economists, taxes are often seen a policy lever to generate government revenue as well as to create incentives. Accountants have taken a similar approach, with an additional interest stemming from understanding taxes as an important item in firms' financial statements for investors to consider for valuation as well as an important item for managers and tax accountants to prepare and submit to the IRS (both firms and individuals; i.e. tax planning).

The literature in economics has mostly seen taxpayers as utility maximizers who are concerned solely with advancing their selfish economic interests. Accordingly, non-compliance can only be curbed through vigilant monitoring and the threat of sanctions and penalties, in line with the classic Becker (1968) model of criminal behavior. As a result, effective enforcement procedures are the routes to success for the authorities.

The literature in psychology sees taxpayers as 'cooperative citizens who are willing to comply if they understand tax laws and perceive the law and the procedures of taxpaying to be fair' (Kirchler, 2007). This cooperation may stem from various sources – social and personal norms, procedural and exchange fairness – which may be leveraged in order to increase compliance. While taxpayers may comply voluntarily if given respect and support, a harsh deterrence approach may backfire and make voluntary compliance less likely in the future (Hessing et al., 1992).

Tax compliance (and the related avoidance and evasion) research within the accounting domain has historically not focused on individual behavior in the same way as economics and psychology. Instead, existing literature focuses on firm-level behavior. Hasan et al. (2017) document a correlation between social capital, measured at the U.S. county level, and firm tax avoidance. Atwood et al. (2012) show correlations between country-level tax system characteristics and firm tax avoidance. Hope, Ma, and Thomas (2013) use geographic earnings disclosures to examine the relationship between financial reporting behavior and tax avoidance. In general, this scratches the surface on different characteristics that correlate with firm-level tax avoidance behavior (e.g. Hoi,

[2] There are signs that this situation is changing. The last few years have seen the emergence of a set of natural field experiments in tax compliance. See, e.g., Hallsworth (2014), Kleven (2011), Ariel (2012), Pomeranz (2015), Castro & Scartascini (2015) and the citations in Floyd and List (2016).

Wu, and Zhang 2013; Graham et al. 2014; Gallemore and Labro 2015; McGuire, Omer, and Wang 2012). Echoing the discussion at the beginning of this section, the overall takeaway is that much of the literature consists of association studies trying to capture forces that incentivize or predict tax avoidance.

Hanlon and Heitzman (2010) reviews the accounting literature on taxation and describes how corporate tax avoidance is multi-faceted and empirically difficult to measure, especially since researchers don't have access to tax return data and must therefore use financial statement data. The main point of their discussion is that there are many different measures of tax avoidance- each of which have their strengths and weaknesses. Much of the discussion concerning prior research in tax compliance revolves around what can be learned given the specific measure of tax compliance or avoidance used in the study.

3. EXPERIMENT I

3.1 Research setting

The field setting was the official process to request payment of UK income tax debts. Most UK taxpayers are not required to submit a tax return, since the great majority of income tax is collected by employers through payrolls. However, around ten million individuals in the UK have to declare their liabilities by submitting an annual Self-Assessment tax return, mainly because they are self-employed or have multiple sources of income.[3] If taxpayers do not pay the correct amount on time, the tax authority has to collect the debt. To do so, an initial statement of account is issued, followed by letters and telephone calls requesting payment.[4]

The natural field experiment concerns the messages in letters sent to Self-Assessment taxpayers who had not made the correct payment by January 31, 2012, and who had not responded to the initial reminder statement. All taxpayers had a debt of between £250 and £100,000 on February 1, 2012.[5] Taxpayers with additional outstanding Self-Assessment debts were excluded, since their situation was more complicated and would have introduced more noise into the results. These procedures resulted in a sample of 105,379 individuals from England, Wales, and Northern Ireland.

3.2 Sampling and randomization

[3] This system requires most of those who submit a return to make two tax payments a year – one by January 31, and one by July 31. http://www.hmrc.gov.uk/sa/need-tax-return.htm

[4] If necessary, the tax authority can enforce payments by seizing and auctioning goods and assets, or taking court action (Her Majesty's Revenue and Customs, 2010).

[5] Debts below £250 and above £100,000 were subject to different actions and therefore could not be included.

The sample of 105,379 taxpayers was divided into fourteen treatment groups, resulting in a mean group size of 7,527. Administrative policy required letters to be sent to all individuals who were late paying their taxes, so the control group was part of this allocation scheme. The sample size means we had adequate statistical power to detect an effect equating to an approximately 1.5 percentage points difference in payment rates, an improvement that was considered to have substantive importance from a policy perspective.

Cases were randomized using an equal allocation procedure and no blocking. The randomization procedure was based on the unique taxpayer number that is created by the tax authority through computer-generated randomization syntax. Cases were selected by assigning the fourteen messages to 84 ranges of these taxpayer numbers. This procedure was used because technical constraints prevented messages from being allocated to taxpayer numbers on an individual basis.[6]

In aggregate terms, the ensuing groups were similar in terms of size, total value, mean debt value, gender, and mean taxpayer age (see Table 1). A logistic regression analysis was used to establish whether membership of a treatment group was significantly predicted by age, gender, size of debt, employment status, or use of an accountant. Of these 70 instances, four were found to be significant at the five percent level: one group had higher debts ($p<0.05$), one had a lower proportion of the self-employed ($p<0.01$), and two had younger taxpayers ($p<0.05$) (Table A1).[7] These differences were very small in substantive terms, but we nonetheless control for these covariates in the analysis below. Since these letters were sent through the national postal service, there was a three day delay before taxpayers received them.[8] This period permitted analysis of whether payment rates varied in the absence of any treatment: they did not.[9] This lack of variation in the pre-treatment window provides another check that the randomization was robust.

The timing of letters was another factor to be managed. The volume of letters meant that they had to be issued over six sequential days and therefore day of issue could present a confounding factor. To prevent this from happening, a Latin Squares design was used to ensure each day received an equal allocation of taxpayer number ranges (see Table 2).

3.3 Treatments

[6] Since each number was generated separately, selecting cases using ranges in this manner was not considered to threaten the integrity of the randomization.
[7] Again, it was not possible to re-randomise within the business constraints of the tax authority.
[8] Letters were addressed solely to the taxpayer with the debt, and all were issued in standard envelopes. Strict taxpayer confidentiality laws meant that recipients could not identify who else was receiving letters unless someone else chose to disclose this information, which means there are limited concerns about spill over effects. The letter issue period did not coincide with any exogenous promotional campaigns, so there is no reason to suspect contamination of the results.
[9] There were no differences by treatment on payment in the first three days.

All letters included basic information on how much was owed and how to pay. The trial letters also featured a short message, in bold typeface, after the first sentence (see Table 1 for the phrases, and Appendix for the control letter). These messages aimed to persuade the recipient to pay their outstanding income tax.

Our primary focus is on messages concerning the oversight behavior of the tax authority. Due to the administrative policy present at the time, there were relatively few meaningful sanctions that were appropriate to invoke in letters.[10] However, the tax authority had recently invested to improve its debt data analytics system to provide day-by-day updates on every late-paying business and individual in the UK (Her Majesty's Revenue and Customs, 2013). Non-payment of debt was thus much more visible to the authorities. It has been shown that greater oversight of tax behavior provided by third-party information is very effective at reducing tax non-compliance, and that there is a causal relationship between visibility and tax compliance in general (Kleven et al., 2011; Sandmo, 2005; Internal Revenue Service, 2006; Bloomquist, 2003; Kagan, 1989). Therefore, informing individuals of this development seemed both appropriate and potentially effective. The message used to communicate this oversight was "*We will be checking how long it takes you to respond to this letter*" – the **Monitoring** group. In order to provide a more robust estimate of this monitoring effect, we also included a variant wording of the same idea: "*We will be checking our records every day to see if you have paid*" – the **Monitoring Day** group.

We include a variety of "control" messages that follow prior studies. The first test message was created to provide a simple reminder that the tax was overdue, and to suggest that the non-payment may have been unintentional. It read: "*Have you overlooked this payment? In case you might have done, I'm writing to give you a reminder*" – we call this the **Reminder** group. While this could be seen as an example of procedural fairness (in terms of assuming that the recipient is predisposed to cooperate), the main purpose of this message was to create a control for "novelty".[11]

The second set of messages concerned the morality of paying tax.[12] We do not propose to discuss this literature in depth here. For the purposes of the current study we focus on two types of

[10] This is not to say that sanctions did not exist. For example, anyone receiving this letter will have already recently been charged a fine of 5% of the balance due, since the payment was thirty days late. Further sanctions are also available should a taxpayer continue to fail to pay. See: http://www.hmrc.gov.uk/helpsheets/sa370-notes.pdf; http://www.hmrc.gov.uk/factsheets/ffc1.pdf

[11] The trial messages reported in Hallsworth et al. (2017) all significantly increased payment rates compared to the control. There was a possibility, therefore, that the introduction of *any* phrase to an existing reminder letter would also produce an improvement. In order to control for such novelty effects, this wording was selected to introduce a new phrase but no new information: it merely restates the obvious purpose of the letter.

[12] Starting with the work of Schmölders (1959) in the 1960s (Frank & Kirchler, 2006), evidence supporting the role of moral concerns in tax compliance has been generated from both theoretical (Gordon, 1989; Erard & Feinstein, 1994; Reckers et al., 1994; Alm & Torgler, 2011) and experimental studies (Trivedi et al., 2003, 2005; Bobek & Hatfield, 2003). There are also various attitudinal surveys that identify the existence of "tax morale", although causal effects on compliance behavior are still unclear (Feld & Larsen, 2012; Wenzel, 2005; Roth et al., 1989). As a result, it has been suggested that governments focus on promoting the public's sense that there is a moral duty to pay tax (Grasmick & Scott, 1982; Alm & Torgler, 2011).

moral concerns. The first is the "intrinsic motivation" that paying tax is a moral act that attracts no reward beyond the act itself (Frey, 1997; Braithwaite & Ahmed, 2005; Deci, 1971). We term this the **Moral Duty** group and present it through the phrase "*Paying your tax is the right thing to do.*" The second aspect draws on the notion of fairness and civic duty: we expect others to pay tax, so we should do the same (Orviska & Hudson, 2003). This introduces a more dynamic, or horizontal, notion of morality as responsibility to others, rather than to an ideal behavior.[13] We attempt to represent this **Moral Equity** group concept in the phrase "*Paying your tax is the fair thing to do.*"

The next set of messages concerned the collective benefits of taxation and the consequences of non-payment. One clear benefit from taxation is the provision of public services. Various studies have suggested that emphasizing the link between tax payments and ensuing benefits may increase compliance (Doerrenberg, 2015; Ortega & Sanguinetti, 2013; Carillo, Castro & Scartascini, 2017). We therefore attempted to assert this link through the following phrase: *"According to a 2009 opinion poll, our most valued public services are the NHS, schools, care for the elderly, and the police. Every single tax payment helps to run these services*" – the **Public Services** group. Note that the public services mentioned in this phrase were selected on the basis of an opinion poll (2020 Public Services Trust 2010), which may increase the credibility of the statement.

Tax compliance brings other benefits, in addition to maintaining public services. At the time of the field experiment there was considerable concern over the size of the UK's public debt and deficit (HM Government, 2010). Maximizing the nation's revenue collection would therefore bring an obvious collective benefit by reducing the costs of financing debt in the international markets. Indeed, the link between tax payments and budgetary health may well be seen as clearer and more direct than that between tax payments and public services, since it involves fewer implementation steps. This link was communicated through the phrase "*Every tax payment we receive means the country has to borrow less money – reducing costs for us all*" – the **Costs & Gain** group.

If higher tax compliance benefits the collective, then non-compliance does the opposite. Clearly, it is equally possible to present this outcome – in other words, to introduce a negative (rather than positive) goal frame. There is some evidence that a negative goal frame may have a larger effect on behavior, although the strength of such effects has been questioned (Taylor 1991, Krishnamurthy, Carter and Blair 2001, O'Keefe et al., 2011). Applying this approach resulted in the

[13] Christian and Alm (2014) tested this second concept by priming participants with a variety of quotes on the theme of "treating others as one wishes to be treated", before asking the participant to formulate this concept in their own words. In the experiment that followed, the primed participants declared significantly more of their income. When regression analysis was applied to control for additional factors, the morality condition was shown to increase compliance by 10%; in comparison, an increase in the penalty rate raised compliance by approximately 8%.

phrase "*Every tax payment we do not receive means the country has to borrow more money - increasing costs for us all*" – the **Costs & Loss** group.

A final message based around the UK's fiscal health was also included. This read: "*According to a 2011 opinion survey, people said that one of the best ways to improve the economy was to reduce the public debt and deficit. But to reduce the deficit we need everyone to pay the tax they owe. Please help us achieve this goal by making your payment now*" – the **Deficit** group. This message was created to invoke the injunctive norm that public opinion considered tax payments as a way to improve the economy.[14] Again, the core aspect of the message was making the link between payments and beneficial collective outcomes salient.

Previous research has shown that presenting the progress already made towards an end state increases an individual's motivation to achieve that state, particularly when their commitment is not strong (Wiebenga & Fennis, 2014; Koo & Fishbach, 2008).[15] As explained above, the recipients of these reminder letters had previously filed their income tax returns and thus had completed at least one part of their duties. There is thus a plausible hypothesis that presenting payment as the second part of a task half completed, rather than as a standalone action (as shown in the control letter), would increase the likelihood it was completed. The message created to test this hypothesis was: "*You have already successfully filed your tax return. All you need to do now is call us to pay the amount you said you owe*" – the **Progress** group.

Finally, the filing of a tax return indicates at least some attempt at tax compliance had occurred.[16] Prior research suggests that such attempts should be acknowledged, thanked, or even rewarded, in order to improve the relationship between taxpayers and authorities (Wenzel, 2006; Feld et al., 2006; Murphy 2005). To test the impact of this strategy, a message of thanks and appreciation was added to the phrase above: "*Thank you for successfully filing your 2010-11 tax return: we appreciate this requires effort on your part. All you need to do now is call us to pay the amount you said you owe*" – the **Progress & Thanks** group

Hallsworth et al. (2017) show that referring to the social norm that others pay their tax on time significantly increases payment rates in two field experiments (Tayler & Bloomfield, 2011; Cardinaels & Yin, 2015). Our next set of messages combined the most effective social norm message from Hallsworth et al. (2017) with the monitoring message: "*Nine out of ten people pay their tax on time - you are currently in the small minority of people that have not paid us yet. We will be checking how long it takes you to respond to this letter*" – the **Monitoring & Norms** group.

[14] The opinion survey referred to is Eurobarometer 74.2: Europe 2020, the Financial and Economic Crisis.
[15] For example, Nunes and Dreze (2006) find that participants were more likely to complete a task if it was presented as involving ten steps, with two already completed, than if it was simply presented as eight steps.
[16] Whether these returns contained a full declaration of taxes owed is a separate – and clearly relevant – question.

The moral duty phrase was added to this message in order to produce the final variation: "*Paying your tax is the right thing to do, and nine out of ten people pay their tax on time. You are currently in the small minority of people that have not paid us yet. We will be checking how long it takes you to respond to this letter.*" – the **Moral & Monitoring & Norms** group. Table 1 presents the full range of messages.

3.4 Results

We first assess the impact of the messages. Figure 1 plots the percentage of people per day in the first 23 days who pay their tax in each of the thirteen treatment groups (as well as the control group). The days for which there is no recorded payment are weekends. From visual inspection, one can clearly see that differences emerge from day 9. Table 3 presents the regression outputs for these data: it shows the effect of the trial letters relative to the control letter during the 19-day sample period. The rationale for selecting this period was as follows. The test messages were only included in the first letter of a multi-letter sequence, which means that the most reliable point at which to measure their effects is the last day before any subsequent letters are received. In this case, that was the 19th day after letter issue, taking into account the variation in postal delivery times. We therefore ran a logistic regression with a dependent variable of payments and cleared balances occurring by 19 days.

Regression 1 of Table 3 shows that four of the thirteen messages significantly increased payments at 19 days ($p < 0.001$). Adding data on the taxpayer's age, gender, size of debt, use of an accountant, or recent debt history did not change these estimates.[17] On a base rate of 57.1%, the simple monitoring statement increased payments by five percentage points (representing a relative increase of 12.5%). The variant message that referred to daily checks increased payments by 2.6 percentage points; this effect is significantly lower than the previous message, which suggests that the precise framing of messages can affect their impact. The message that added social norms to this approach had a marginal effect of 4.4 percentage points; adding in the moral duty statement led to a 3.7% increase in payments. Neither of these effects were significantly different from that of the simple monitoring statement.

We also conducted a set of subgroup analyses. Columns III and IV of Table 3 show the results for taxpayers who did or did not incur at least one debt in the last three tax years, respectively. This variable has been found it to be a strong predictor of tax compliance behavior (Hallsworth et al., 2017). Two results are interesting from a theoretical perspective. First, the gentle reminder led

[17] "Recent debt history" was a dummy variable that indicated whether the taxpayer had paid tax late in any of the preceding three tax years.

to a 2.6% points increase in payments for the group who had not been in debt recently, whereas it led to a (non-significant) decrease in payments for the recent debtor group. Second, the moral duty message increased payments by 3.1% points for those without recent debts but led to a (non-significant) decrease in payments for recent debtors.

Tables 4 and 5 show subgroup analyses for the main covariates available in the dataset. Table 4 shows that none of the treatments produced a significant increase in payment rates amongst the sample of female debtors. In some cases, this may be an issue of statistical power (since women make up around a quarter of the total sample), but in other cases letters that increased payments amongst men produced a negative, though not significant, effect amongst women. The differences between the male and female treatment coefficients are meaningful and significant at the one per cent level. The results do not differ greatly according to debt quartiles, whether a taxpayer is below or above the median age, or whether they use an accountant or not.[18]

Table 6 shows the effects of the treatments over a longer time period. While the duration of effects is an important issue, caution is needed when interpreting this table. After the initial letter was sent out (i.e. after the 19th day), the tax authority undertook a range of follow-up activities on a non-randomised basis. Therefore, we can see some changes in the coefficients that might not be expected. However, it is noticeable that the coefficients for the "Monitoring" and "Monitoring & Norms" letters remain positive and meaningful ($p< 0.01$) until the 70-day mark.

Finally, we calculate that £4.7 million of revenue was accelerated in the first 19 days from introducing these test messages. This figure was calculated by taking the marginal effects that are significant at the 0.05 level and multiplying them by the average debt value for the relevant treatment groups and multiplying by the size of the treatment group.

Our moral costs results are seemingly at odds with prior studies, such as Hallsworth et al. (2017), which find that moral costs increase tax compliance. It should be noted that our current moral cost treatments are different to those in Hallsworth et al. (2017). In Hallsworth et al. (2017), the moral cost was phrased as "9 out of 10 people pay their tax on time. You are in the minority of people who do not". This is a very specific moral cost in the sense that it required detailed information on those who were performing the behavior and those that were not. In our current paper, we wanted to generalize the moral cost message so that it can be used in other contexts. For example, our moral duty and moral equity treatment groups did not mention the descriptive social norms, they simply state, "Paying your tax is the right thing to do" and "Paying your tax is the fair

[18] One interesting point to note is that the "Progress" messages appeared to fare worse for those people using accountants. This may be because these messages state that the recipient already filed their tax return: the letter may be received by the accountant, who may not consider that they personally filed the tax return.

thing to do" respectively. The apparent difference in taxpayer response to the moral cost messages between this study and Hallsworth et al. (2017) suggests that when moral costs are presented to taxpayers, the precise framing of the message matters. This finding helps shed light on mixed results concerning moral costs in the tax compliance literature more generally (e.g. De Neve et al., 2021).

A final word of caution is warranted concerning the interpretation of our results. Though a significant treatment effect concerning taxpayers' receipt of a particular message provides evidence that the underlying incentive for tax compliance is salient in our setting, the lack of a treatment effect does not mean that the underlying incentive does not matter. For example, it could be the case that moral concerns are important in our setting, but individuals are already acting in accordance with their morals and thus an additional nudge is ineffective at changing behavior.

4. EXPERIMENT II

4.1 Research setting

Our second natural field experiment retained exactly the same setting as the first but was implemented a year later. The objective of this second experiment was to further examine which kinds of monitoring and oversight costs were salient to taxpayers given their importance in the first experiment. As before, letters were sent to Self-Assessment taxpayers with standalone debts of between £250 and £50,000 on February 1, 2013, and who had not resolved matters in response to the initial reminder statement. We obtained 204,936 individuals in our sample from England, Wales, and Northern Ireland.

4.2 Sampling and randomization

The total sample of 204,936 was randomized into twelve groups, with a mean sample size of 17,078. Again, technical limitations prevented blocking on background variables prior to the letter issue. To further reduce the possibility of imbalanced samples, we extended the randomization procedure by selecting 156 ranges of taxpayer reference numbers. Computer-based randomization was then used to assign these ranges to one of the twelve groups and one of the seven possible issue days (replacing the Latin Squares design used to allocate ranges to days in the first experiment).[19] The ensuing groups were similar in size, total debt value and debt value, mean taxpayer age, and

[19] To clarify: there was freedom for any number range to be assigned to any letter and any day of issue. Two ranges were assigned to each letter for each of the first six days, and one range to each letter for the seventh day, in order to align with the tax authority's business procedures.

gender ratios (Table A2).[20] The procedure also resulted in even allocation of letters across the seven issue days (Table 8).

4.3 Treatments

The structure of the letters remained the same as in the first experiment. All letters contained basic information about the debt and how to pay it. Again, the test messages were included after the first sentence, and all other aspects of the letters were identical to the previous year.

The trial focused on two main factors: the kind of costs taxpayers may incur, and the timing of these costs. Experiment I demonstrated the effectiveness of stating that the tax authority would take action in response to the recipient's behavior, with the implicit costs that accompany such action. Experiment II therefore explores what kinds of these costs and consequences are salient to taxpayers. More specifically, three types of costs are examined: (i) the financial cost from interest charges; (ii) the time and effort costs of further interactions with the tax authority; and (iii) the non-specific costs from enforcement action.

The *timing* of these costs was introduced in order to investigate the possibility that procrastination plays a role in the tax payment decision. A crucial aspect of procrastination is present-based or quasi-hyperbolic time preferences, which lead the individual to overweight immediate costs compared to future costs (O'Donoghue and Rabin 1999). Since it was not possible to randomize the actual time at which costs were incurred, the way these costs were presented was randomized instead. This was done by either presenting the immediate gains achieved by paying now (present-frame), or by presenting the increased future losses to be incurred by paying later (future-frame). If time preferences were present-biased, then the present-frame is likely to result in higher payment rates than the future-frame.[21]

These two dimensions of costs and timing of costs were used to generate the treatment messages. The first type of cost was interest charged on the debt, the mention of which had previously been shown to raise tax payment rates by four percentage points (Hallsworth et al., 2017). The present-frame of this cost was "*We are charging you interest daily. You will save money if you pay now*" – the **Interest Today** group. The future-frame was "*We are charging you interest daily. You will lose money if you pay later*" – the **Interest Tomorrow** group. The second type of cost was

[20] As for Experiment I, a logistic regression was run to analyse whether any covariates significantly predicted assignment to a particular group. Size of debt and gender were not significant predictors for membership of any group; the "Interest & Effort" group were less likely to be self-employed and use an accountant ($p < 0.05$). We do find that age was a significant predictor for five of the groups: "Costs Further Action", "Contact Now" (both $p < 0.05$), "Less Effort", "Contact Now Online", and "Interest & Effort" ($p < 0.01$).

[21] There are, of course, other conditions that would need to be fulfilled: the framing itself would need to be effective, and the cost itself would need to be salient.

the time and effort required to deal with future communications from the tax authority. The present-frame read "*Paying now means less hassle because we won't contact you about this again*" – the **Less Effort** group. The future-frame was "*Paying later means more hassle because we will contact you about this again*" – the **More Effort** group.

The third and final type of cost was the non-specific cost of future debt pursuit or enforcement action. The present-frame of this cost was "*If you pay this debt now, you can stop us taking any further action*" – the **Costs Further Action** group. The detail of the future action was not specified at this stage because there were a variety of options open to the tax authority, which would be set out in a later communication. There was no future-frame for this cost, since that would have involved stating that the individual would have no ability to stop further action later – this was not in line with the tax authority's policy of resolving debts without enforcement wherever possible. Instead, the opportunity was taken to compare the effect of mentioning enforcement costs versus a cognitive cost to the individual. This cognitive cost was presented as the concern brought about by a failure to address the situation: "*If you pay this debt now, you won't have to think about it anymore*" – the **Costs Peace** group. Note that the message structure was identical to the previous message, allowing a better comparison of the two types of cost.

The next step was to combine different types of costs. The three types of cost were all given different message structures because this allowed them to be combined without creating undue repetition. However, we did not include all 24 combinations of messages, since this would have considerably reduced the trial's power. Rather, we retained the present-framed interest charge (since there was empirical and theoretical evidence that this would be effective) and matched it with the present-framed messages related to effort costs and enforcement costs. The **Interest & Effort** group message read: "*We are charging you interest daily. You will save money if you pay now. Paying now means less hassle because we won't contact you about this again.*" The **Interest and Action** group message was: "*We are charging you interest daily. You will save money if you pay now. If you pay this debt now, you can stop us taking any further action.*"

The final set of messages focused on the timing of the action required, rather than the timing of the costs. The intent here was to examine whether taxpayers would respond more to a request for future action or a request for immediate action. There are, of course, practical and theoretical reasons to favor the latter. The default position for the tax authority is to require an instant response, since the tax is overdue. If taxpayers have present-biased time preferences, then they are likely to choose an option to respond later, but then postpone the action further when the time comes. However, there is also much evidence that making a plan to carry out a future action increases the likelihood that it will be completed. The effectiveness of this "implementation intentions" approach

has been demonstrated for various behaviors, including healthy eating, exercise, and perseverance with tasks (Gollwitzer, 1999; Gollwitzer & Sheeran, 2006). It was therefore included to test its effects in a tax compliance context.

The message developed to represent the request for immediate action was "*Please contact us as soon as you receive this letter*" – the **Contact Now** group. When developing the future framing of this "Contact Now" message, it was judged that the point in time that would clearly be perceived as the 'future', yet which would not introduce unnecessary delays to payment, was the following day. However, to simply present the phrase "Please choose a time tomorrow to contact us" would have been problematic, since recipients would have lacked the necessary information about what times were viable. On the other hand, to simply introduce a future-framed message with this information would risk confounding the specific effect of the future frame. These issues were addressed by creating two new messages. The **Contact Now Information** group added details to the statement above to produce a present-framed message with additional information: "*Please contact us as soon as you receive this letter. We are open Monday to Saturday 8am to 8pm, Sundays 8am to 5pm - or you can pay online any time.*" The **Contact Later Information** group then introduced the future frame: "*Please choose a time tomorrow to contact us. We are open Monday to Saturday 8am to 8pm, Sundays 8am to 5pm - or you can pay online any time.*" A comparison of these two messages will therefore isolate the specific effect of the future frame. All messages are summarized in Table 7.

4.4 Results for Experiment Two

As with Experiment I, the main dependent variable is whether the taxpayer makes a payment. Figure 2 plots the percentage of people per day in the first 48 days who pay their tax. In this experiment, the tax authority's procedures meant that the latest point before any further communications were received was 22 days after the issue of the first letter. Therefore, we analyze whether payment had occurred by the end of the 22nd day. The logit model used was identical to the one presented above, except that it included eleven treatment dummy variables. The covariates included were the same as for Experiment I: age, gender, size of debt, self-employed status, and use of an accountant. Table 9 gives the outputs from the regression analysis. A comparison of columns (I) and (II) shows that the coefficients for the letters did not change after the covariates were added to the model.

We first examine the treatments representing different kinds of costs. As regression (I) in Table 9 demonstrates, the messages concerning interest charges and time costs did not significantly differ from the control at the ten percent significance level. The treatment that warned of

enforcement costs ("Costs Further Action" group) increased payment rates by 4.5 percentage points ($p < 0.01$). However, the peace of mind cost message – which had an identical structure – did not differ significantly from the control. This suggests that focusing on enforcement costs is more effective than peace of mind costs. Turning to the combined treatments, we can see that the "Interest & Effort" treatment did not significantly increase payment rates, while payments in the "Interest & Action" group were 1.7 percentage points higher than the control ($p = 0.002$).

The messages based around the timing of the action produced interesting results. The "Contact Now" message increased payment rates by 2.2 percentage points ($p < 0.001$). However, as the section above explains, the effect of the present and future framing can only be isolated by comparing the two longer messages that both have information about opening times. The present-framed version of this message did not significantly increase payment rates at the 0.05 significance level, but its future-framed equivalent raised payment rates by 2.9 percentage points ($p < 0.001$). This provides tentative evidence that an "implementation intentions" approach could be used to help people meet their tax obligations.

Finally, the results above also suggest that the length of messages may affect their impact. The marginal effect of the non-specific costs message was reduced from 4.5 percentage points to 1.7 points (a significantly lower level, $p<0.1$) when the warning about interest costs was added, even though this warning did not reduce payments when presented on its own. Similarly, the 2.2 percentage points improvement from the "Contact Now" message was eliminated when further information was added. There is thus some (very limited) evidence that making messages longer may reduce their impact.

Regression (II) controls for the background independent variables of the individuals in our sample. It is clear that the coefficients do not change once these independent variables are included. It is also worth considering the covariates included in this table. We can see that every year of age increases the payment rate by 0.4% points (and reduces the time to pay by 0.17 days), a finding that is in line with previous studies. Similarly, the payment rate for men is 3.2% points lower than that for women (or 1.48 days), which also accords with existing studies. We also note that having a debt in any of the previous tax years led to a 4.8% point (1.33 days) reduction in payment rates (see also Hallsworth et al. 2017). Interestingly, we observe that the use of an accountant had no significant effect on payment rates, in contrast to Experiment I and other previous studies (Erard, 1993); moreover, being self-employed appeared to *increase* compliance by 12% points, which is the opposite effect to that predicted by existing studies.

Regressions (III) and (IV) separate out the results according to whether the taxpayer had been late in one of the preceding three tax years. The results are similar, although it appears that

interest messages are effective amongst those without recent debts, and not effective amongst those who have. Tables 10 and 11 give subgroup analyses of the main outcome measure (payment at 23 days) by each of the covariates mentioned above. Two points are particularly worthy of note. First, the "Interest Today" message appears to be more effective for taxpayers above the median age (1.8% points increase in payment rate, p=0.020) than for those below the median age (1.7% points decrease in payment rate, p=0.028); a similar result can be seen for the "Interest Tomorrow" group in regression (III) in Table 10. Second, we can see that the peace of mind message may backfire for male taxpayers (1.2% points decrease in payment rate, p = 0.049) and for those below the median age (2.1% points decrease, p=0.005). The "Costs Further Action" treatment group was effective across all subgroups.

Table 12 gives the results for the treatment groups at the end of different phases in the debt collection sequence. As before, we were not able to maintain the randomization in these later stages, and thus it is problematic to make causal claims from these data. With this in mind, we can see that the "Interest Tomorrow", "Less Effort" and "Contact Now Information" groups appear to have significantly lower payment rates over both 48 and 70 days.

We have made a provisional calculation of the revenue accelerated from this trial, using the same approach as for Experiment I. We calculate the added revenue in the first 23 days by taking the coefficient for each message and multiplying it by the number of people in each treatment group, then multiplying that by the average debt. This produces an estimate of £5.16 million advanced. When considering these sums, it should be noted that the costs of this intervention were very small.

5. EXPERIMENT III

5.1 Research Setting

Our third field experiment was implemented in 2021. The objective of the third experiment was to further examine why the costs of further action treatment was salient to taxpayers given its importance in the second experiment. The structure of the third experiment kept many aspects of the first two field experiments but also incorporated many changes. Because the tax authority would not allow messages that stated specific enforcement actions, we instead conducted an experiment with 7,064 Prolific Academic members[22], who completed this study in exchange for a $1 remuneration. 4,236 (59.97%) of the sample were female, 2,732 (38.67%) were male, 48 (0.68%)

[22] 7,066 Prolific Academic members participated in the study. Two participants completed the study twice; their responses were excluded from the analyses.

specified their gender as 'non-binary', while 48 (0.68%) preferred not to disclose their gender. The mean age of respondents was 36.78 years.

Only participants between the ages of 21 and 65 and residing in the UK were included in the experiment. In addition, we screened the participants based on their response to the following question: "Have you engaged in entrepreneurship/run your own business?" We only included participants who responded either "I am currently doing this" or "I have in the past"[23]. Because the experiment was conducted through Prolific, we did not conduct a natural field experiment. Instead, repayment rates were purely hypothetical based on a willingness to pay scale.

The experiment asked participants to rate their willingness to pay an outstanding tax obligation on a 11-point Likert scale, where 0 = 'Certain not to pay', and 10 = 'Certain to pay' (see Appendix Figure A1). Messages were incorporated before they were asked their willingness to pay in order to measure treatment responses. Using 11-point Likert scale, we also asked participants "How severe do you think the punishment would be if you do not respond to the letter?" (0 = 'No action', and 10 = 'Maximum penalty possible') and "In your opinion, how realistic was the punishment mentioned in the letter?" (0 = 'Not at all realistic', and 10 = 'Very realistic').

To assess participants' risk propensity across multiple domains, we administered Dohmen et al.'s (2011) General Risk Question. Unlike the Domain-Specific Risk-Taking Scale (Blais & Weber, 2006), which taps into specific domains of risk-taking over 40 different items, the General Risk Question is thought to aggregate these domains into one simple question. Participants self-reported their response to the question "How willing are you to take risks, in general?" using a Likert scale ranging from 0, which indicated not at all willing to take risks, to 10, indicating high willingness to take risks. The General Risk Question's simplicity allows for a sensitive measure of one's risk risk-taking attitude, and thus we used this as a proxy for participants' risk propensity. Literature measuring psychometric properties revealed the General Risk Question to be an accurate measure of risk propensity across a variety of domains (Dohmen et al., 2011), with high test-retest reliability (Lönnqvist et al., 2015).

5.2 Sampling and Randomization

Participants were randomly assigned into one of 7 groups corresponding to the 7 different messages that were tested in the experiment (see Table 13). For this we used the randomization algorithm enabled by Qualtrics Experience Management Platform. Each group consisted of approximately 1,000 participants (ranging from 1,005 to 1,014).

[23] In the hypothetical scenario provided to the participants, they were asked to imagine that they own a cash constrained business.

5.3 Treatments

Due to sample size considerations, we only tested 7 different messages. Though this number is significantly smaller than our previous field experiments, the relatively narrow focus of the third experiment only required a smaller number of messages.

Our first message was an identical copy to the "Costs Further Action" message in experiment two. This message was included as a benchmark so that we could compare its effectiveness to additional messages.

The next group of messages concerned the possibility of the taxpayer receiving a fine for not paying their taxes. The motivation behind including these messages was the possibility that the "Costs Further Action" message was producing a treatment effect in experiment two because it was proxying for an enforcement action that taxpayers likely anticipated – being fined.

However, it was unclear ex-ante whether the specific nature of the fine mattered. To address this, we included three different variations of a message concerning fines. The first message **Fine Non-specific** stated that the taxpayer could receive a fine but did not mention the exact amount of the fine. The second message **Fine Average** was similar to the first message, but instead stated that the fine was around the average fine received for noncompliance in our setting. Finally, **Fine Maximum** repeated the second message but changed the size of the fine to the maximum fine that could be levied in our setting.

The final group of messages were similar in nature to the messages about fines, but this time the messages made reference to jailtime. Once again, we included messages about non-specific jailtime (**Jailtime Non-specific**), average jailtime (**Jailtime Average**), and maximum jailtime (**Jailtime Maximum**). We present all treatment messages and descriptive statistics in Table 13. In addition, Figure 3 presents a box plot of participants' willingness to pay an outstanding tax obligation (i.e. repay likelihood) by treatment group.

5.4 Results for Experiment Three

Table 14 Column I presents OLS results with "Costs Further Action" as the control group. When we compare the benchmark "Costs Further Action" message with the group of fine messages, we do not find a significant difference in likelihood of tax repayment from the control for "Fine Non-specific" or "Fine Average". However, when the size of the fine is increased to the maximum (i.e. the fine is more than tax owed), this increases the likelihood of repayment when compared to the "Costs Further Action" treatment ($p<0.01$). The "Fine Maximum" message also appears to

produce a larger treatment effect than both the average fine ($p < 0.01$) and non-specific fine ($p < 0.05$) messages.

Next, when we compare the "Costs Further Action" message against the group of jailtime messages. When testing for each type of jailtime messages, messages mentioning jailtime produce a significant increase in the likelihood of repayment compared with the "Costs Further Action" message: for non-specific jailtime the coefficient is .203 ($p<0.05$); for average jailtime the coefficient is .237 ($p<0.01$). However, the maximum jailtime does not produce a significantly larger treatment effect compared with the "Costs Further Action" message. These results do not change significantly when controlling for age, gender, and risk propensity (Column II). Column III extends the analysis by including information in which we asked participants how realistic they perceived the punishment and how severe they thought the punishment would be if the letter from the tax authority was not responded to. When realism and punishment severity are added as controls, the "Jailtime Maximum" message becomes significantly positive ($p<0.001$). Considering objectively high punishment severity communicated through the "Jailtime Maximum" message, Column III results suggest that the "Jailtime Maximum" message is not always seen as realistic, which dampens the effect of the punishment. Adding age, gender and risk propensity as additional controls in Column IV does not significantly change the results reported in Column III.

To understand why jailtime produces stronger repayment rates than fines ("Fine Non-specific", "Fine Average"; Figure 3), we make use of information in which we asked participants to tell us how severe they thought the punishment was. We thus examine a mediation model which includes perceived punishment severity as a mediator of the relationship between treatment and likelihood of tax repayment. We use the SPSS PROCESS macro (Hayes, 2017) to test whether perceived punishment severity mediated the direct effect of treatment[24] on willingness to pay an outstanding tax obligation. In a separate model block we also tested whether perceived realism of the messages mediated the direct effect of treatment on willingness to pay. The significance of the direct effects (i.e. the effect of the predictor variable on the outcome variable) and indirect effects (i.e. the effect of the predictor variable on the outcome variable via the mediator) was evaluated by means of 5,000 bootstrap samples to create bias-corrected confidence intervals (CI's). The fact that the bias-corrected CI's do not cross over zero means that there is likely to be a genuine indirect (mediation) effect (Hayes and Preacher, 2014).

[24] In each model reported in Table 15 the independent variable was manipulated at two levels: a message of interest vs. control ("Costs Further Action").

Mediation analysis reported in Table 15 suggests that stating that the participants could face jailtime, compared to control ("Costs Further Action"), increases their likelihood of repayment significantly through the punishment severity channel. The analysis reveals a significant indirect effect of treatment (manipulated at two levels: message of interest vs. control) on likelihood of repayment through punishment severity, for such messages as "Jailtime Non-specific" (b = 0.088, 95% CI [0.055, 0.123]), "Jailtime Average" (b = 0.036, 95% CI [0.019, 0.054]), or "Jailtime Maximum" (b = 0.019, 95% CI [0.009, 0.0529]), compared to control ("Costs Further Action"). Similar findings apply to the "Fine Maximum" vs "Costs Further Action" treatments (b = 0.095, 95% CI [0.077, 0.115]). Notably, while the significance of the mediation effect holds, the direction of the indirect effect reverses when considering "Fine Average" vs "Costs Further Action" treatments (b = - 0.046, 95% CI [-0.068, -0.026]).

The results reported in Table 15 also reveal the mediating role of perceived realism on the relationship between treatments and the likelihood of repayment – for all three jailtime messages and for "Fine Non-specific". Notably, compared to "Costs Further Action", all three jailtime messages are seen by participants as less realistic which produces a negative indirect effect of treatment on the likelihood of repayment through perceived realism: "Jailtime Non-specific" (b = - 0.077, 95% CI [-0.110, -0.146]), "Jailtime Average" (b = -0.070, 95% CI [-0.088, - 0.054]), "Jailtime Maximum" (b = - 0.054, 95% CI [-0.066, -0.043]). "Fine Non-specific" is considered to be more realistic than "Costs Further Action", which explains the positive indirect effect reported in Table 15 (b = 0.291, 95% CI [0.210, 0.375]).

Overall, it appears as though the "Costs Further Action" message produces meaningful treatment effects in experiment two because taxpayers 'fill in the blank' that the further enforcement action is likely a fine. Interestingly, stating the specific amount of the fine does not affect payment rates unless it is severe (in our experiment- the maximum). This result is potentially important for policymakers because it suggests that regulators can produce meaningful changes in behavior with messages without necessarily having to commit to specific enforcement actions or magnitudes.

That being said, the threat of prison is particularly salient. If tax authorities have the latitude to send messages with explicit threats to prison, messages about jailtime could produce even more pronounced changes in payment rates (Holz et al. 2020). Once again, the severe jailtime message only does as well as the less severe punishment messages (i.e., "fine non-specific", "fine average" and "costs further action")- indicating that severe punishments that appear to be unrealistic for participants do not change payment rates. It would be interesting to see if these results translate to the field.

6. EXPERIMENT IV

6.1 Research Setting

Our fourth experiment was also implemented in 2021. The objective of the fourth experiment was to reinforce the results from experiment three using a complementary empirical approach in a controlled setting, while also measuring participants' visual attention to the presented messages and induced psychophysiological arousal. The structure of the fourth experiment kept many aspects of the third experiment but was conducted in the lab. Undergraduate and graduate students from a large Australian university were invited to participate in this study and were given a similar task to those in the Prolific study. As an incentive, students received bonus course credit for participation.

To be included in the experiment, participants had to have experience in paying taxes. Because the experiment was conducted in person during COVID[25], our sample size was limited. We were able to recruit a total of 116 participants. As in experiment three, repayment rates were purely hypothetical based on a willingness to pay scale. Using an eye tracking device and an electrodermal activity assessment device, two alternative measures of psychophysiological arousal[26] which are commonly used as biomarkers of fear, stress, and anxiety (Hyde, Ryan, and Waters, 2019; Sirois and Brisson, 2014 ; Laeng and Alnaes, 2019) were collected[27]. Both measures are closely associated with the activity of the sympathetic nervous system, a branch of autonomic nervous system which is involved in the regulation of unconscious bodily functions. In addition, we collected and analyzed participants' eye tracking (gaze fixation) data to control for their attention across treatments while they were reading the letter from the tax authority.

It is a well reported fact that pupil diameter increases when respondents process emotionally or cognitively engaging stimuli (Bradley, Miccoli, Escrig, and Lang, 2008; Henderson, Bradley, and Lang, 2018; Sirois and Brisson, 2014). From the neurophysiological perspective, the change in the pupil diameter occurs due to the activation of the sympathetic nervous system or the inhibition of the parasympathetic nervous system (Unsworth and Robison, 2017). Pupillary response has been used as a biomarker for cognitive and emotional arousal in decision-making tasks in the accounting (Rose, Rose, Rotaru, Sanderson, and Thibodeau 2021, 2022; Brink, Gouldman, Rose, Rotaru 2020)

[25] The experiment was conducted in between 5th and 6th lockdown in the city of Melbourne (Australia) so the number of participants was limited due to the restrictions associated with COVID pandemic.
[26] Arousal is defined as a change in a person's "physiological and/or psychological responsiveness to internal or external stimuli" (Howells, Stein, and Russell, 2010, 39).
[27] The use of biomarkers of stress and arousal to support theory testing follows an emerging research trend in the accounting literature (e.g. ; Brink, Gouldman, Rose, Rotaru 2020; Cardinaels and Feichter, 2021; Rose et al., 2021, 2022).

and the economics literature (Alós-Ferrer, Jaudas and Ritschel 2021; Wang, Spezio and Camerer 2010).

We use an eye tracking device to record the change in diameter of participants' pupils. Our measure of *Pupillary Response* has been established in accounting literature (Rose et al., 2021, 2022; Brink et al., 2020). It represents a continuous proportional variable, which reflects the individual-specific percentage change between: (i) the average pupil dilation, measured prior to commencement of the main experiment by requiring participants to look at a fixation cross displayed on the screen for 60 seconds (i.e., individual baseline; see Appendix Figure A2); and (ii) the average pupil dilation while participants attended to experimental stimuli (i.e., when they were reading the letter from the tax authority). This approach to measuring pupillary response, known as the divisive method (i.e., pupillary response is calculated as a [pupil size – baseline] / baseline), aligns with reported best practice (e.g., Hayes and Petrov 2016, Brink et al., 2020; Rose et al., 2021, 2022). Specifically, in this study we include participants' pupillary responses when their attention is fixated on the letter from the tax authority, which is achieved by synchronously collecting participants' eye tracking (gaze fixation) data and pupillary data (as reported in Brink et al., 2020; Rose et al., 2021, 2022).

In addition to change in pupil diameter, we collected an alternative measure of arousal which tracks the changes in sweat gland activity controlled by the sympathetic (i.e. arousing) nervous system. For this, we measured participants' electrodermal activity, i.e. the electrical conductivity of their skin (Boucsein, 2012). The change in electrodermal activity can be caused by emotional states such as anger, fear, and anxiety (Bach, 2016; Boucsein, 2012). It may reflect the conscious expectancy of a likely negative outcome but can also indicate an unconscious emotional process thus representing "an objective, non-voluntary indicator of anticipation and arousal" (Hyde, Ryan, and Waters, 2019, 2). In fact, there is evidence suggesting that the changes in pupil size and skin conductance may covary in response to emotional stimuli such as affective pictures (Bradley et al., 2008).

Our alternative measure of arousal, *Skin Conductance Response (SCR)* was calculated using the same divisive method used to calculate pupillary response: for each participant we calculate individual baseline while participants are asked to stare at the fixation cross for 60 seconds at the start of the experiment. Using synchronously captured eye tracking data, we extract participants'

skin conductance level[28] (in MicroSiemens[29]) while they are reading the letter from the tax authority and calculate the change in their skin conductance level from baseline ([skin conductance level – baseline]/baseline).

In addition to these two measures of arousal, we used an eye tracking device to acquire two alternative measures of participants' attention while they were reading the messages from the tax authority, total dwell time (fixation duration), and fixation count in relation to those messages (Chen, Jermias, and Panggabean, 2016; Fehrenbacher, Schulz, and Rotaru, 2018; Dalla Via, Perego, and Van Rinsum, 2019; Brink et al., 2020; Rose et al., 2021, 2022; Rotaru, Kalev, Yadav, and Bossaerts 2021).

The gaze and pupillary data were collected using a Tobii TX300 eye tracking system, with a sampling rate of 300 Hz, connected to a 23-inch monitor, with a screen resolution of 1920x1080 pixels (see Appendix Table A3 for additional information). Participants' electrodermal activity data while they were reading and processing the letter from the tax authority was collected using a Shimmer3 GSR+ device, with a sampling rate of 128Hz, using Consensys multi sensor management software as well as Tobii Pro Lab (version 1.171.1) software. The latter (installed on six TobiiTX300 devices used in this study) allowed us tosynchronously capture and analyse eye tracking, pupillometry and skin conductance data around the selected area of interest, which was structured around the letter from the tax authority presented to participants on the screen of the eye tracking device.

Since pupillometry data collection requires the minimum presence of light to reduce the effect of the natural pupillary light reflex (Sirois and Brisson, 2014), the experiment was conducted in a dimly lit booth. Before undertaking the main part of the experiment, participants were asked to remain still for a duration of 60s while a blank grey screen[30] with crosshairs (fixation cross) in the center was presented to them. We use the period of 60s when the participants were instructed to stare at the fixation cross in the beginning of the study as the baseline period for calculating the percentage change in the individual's psychophysiological arousal (indexed by pupillary or skin conductance response) when attending to the experimental treatment (see Appendix Figure A2).

[28] For our analysis we only used skin conductance data captured while participants' eyes were fixated within the area of interest, which was framed around the letter from the tax authority. This assured that we are capturing participants' arousal directly associated with the content of the letter and not any other cues presented on the page or outside the screen of the eye tracking device.

[29] µSiemens, a unit of electric conductance: 1 Siemens = 1/Ω (where Ω denotes Ohm), so 1 µSiemens = 1/MΩ (Boucsein, 2012).

[30] Reduced luminance of the screen was used to reduce the effect of pupillary light reflex at the baseline (Laeng and Alnaes, 2019).

6.2 Sampling and Randomization

Participants were randomly assigned (using a randomization algorithm from https://www.random.org/) into one of 7 groups corresponding to the 7 different messages that were tested in the experiment. Each group consisted of approximately 17 participants.

6.3 Treatments

To ensure consistency across experiment three and experiment four, we utilize the exact same messages as in experiment three.

6.4 Results for Experiment Four

Table 18 presents our baseline results as in the prior experiment. Similar to experiment three, the analysis suggests that there is no significant difference in payment rates between "Costs Further Action" and the fine group treatments (Column III). Also, in line with the experiment three results, we find a marginally significant difference in likelihood of repayment between "Costs Further Action" and jailtime group treatments ($p< 0.05$) (Column III). At the individual message level in Column I, the "Jailtime Average" message is significant. The "Jailtime Non-specific" message is only significant when we add controls for age, gender, and risk propensity in Column II. Interestingly, the "Fine Maximum" message is never significant even though we found a significant treatment effect in experiment three. However, upon closer inspection the t-stat in both of the prior two instances is barely shy of 1.96 (1.93), which indicates the lack of significance is likely due to lower statistical power.

Along with behavioral data (Table 17), we collected eye tracking, pupillometry, and electrodermal activity data while the participants were reading the letter during the experiment (Table 18). Pupillary response and SCR levels across treatments are depicted in in Figures 5 and 6. ANOVAs using eye tracking data reveal the lack of between-group difference for either of the two collected eye tracking measures ($p>0.05$, not tabulated), indicating a comparable cognitive effort exhibited by participants when reading and processing the messages. ANOVAs of the arousal data show significant between-group difference, both for pupillary response ($p=0.03$) and SCR ($p<0.001$, not tabulated).

Upon conducting mediation analysis using SPSS PROCESS macro, we find a significant indirect effect of treatment groups ("Costs Further Action", fine messages, and jailtime messages[31])

[31] Given a low sample size, we clustered those groups.

on repayment likelihood through arousal, as proxied via pupillary response (b = 0.104, 95% CI [0.041, 0.245]), as well as via SCR (b = 0.247, 95% CI [0.100, 0.430]). Given there was no between-group difference in the level of attention when participants were reading the letter from the tax authority, the observed arousal was not triggered by their attentional effort. Instead, it reflected the activation of participants' sympathetic nervous system due to their emotional response to the letter from the tax authority, as proxied by participants' pupillary response and SCR.

Overall the results of the lab experiment are consistent with the Prolific study. The "Costs Further Action" message does about as well as messages concerning non-specific or average fines. Once again, messages concerning jailtime are typically more effective than the rest of the messages.

7. DISCUSSION

Overall, the results from the first natural field experiment suggest that messages that made salient the degree of oversight being exerted by the tax authority significantly increased payments; whereas messages representing moral concerns, public goods, and a supportive approach did not increase payment rates within three weeks of the letter issue.

Our second natural field experiment provided further exploration by examining what kind of costs were most effective in increasing tax compliance. The results suggest that not all kinds of costs are effective at changing behavior. Messages based on interest charges, time costs, and peace of mind costs did not significantly increase payment rates for the sample as a whole, at least in the formulations used here. In contrast, messages warning of further action (i.e. non-specific costs) raised payment rates between 1.7 and 4.5 percentage points. It is clear that the legal powers of the tax authority – even if not made explicit to the taxpayer – are an effective route to increasing compliance.

Our third experiment used a sample of participants through Prolific to further disentangle why messages concerning further action were salient in experiment two. The results suggest that the "Costs Further Action" is about as effective as stating either a non-specific or average fine. However, the results also show that if the tax authority stated a severe fine or jailtime, then payment rates could increase even further. Nevertheless, payment rates for the "Costs Further Action" message reinforce that authorities can still achieve meaningful payment rates even if they can't commit to a specific enforcement action.

Finally, experiment four repeats experiment three in the lab. We recruit over 100 students and measure their physiological responses to the same set of messages as in experiment three. First,

we find results that are largely consistent with experiment three. Second, the change in payment rates seem to be driven by the activation of participants' sympathetic nervous system as an emotional response to the letter from the tax authority.

The wider lesson that can be drawn from these experiments is that the application and use of field experiments reveals new opportunities to alter *the way* in which authorities interact with taxpayers. This means that finding new ways to present costs could yield benefits to tax authorities and point towards new routes to compliance for tax authorities without changing the actual audit probabilities or the penalty for not complying.

We close on some remarks regarding generalizability. First, our natural field experiments are conducted in the United Kingdom, which likely differs in many ways from other countries, such as the United States. However, though tax systems differ across countries, our analysis attempts to focus on more fundamental aspects of human behavior, which should be less likely to deviate across settings. This is supported by the findings of the lab experiment which was conducted in Australia.

Second, we focus on individuals. Individuals may behave differently than small businesses and corporations. Firms are still comprised of people that make decisions; therefore, our results may be able to provide some guidance for encouraging tax compliance in these settings. We encourage future research to examine this possibility further. Regardless, individual tax compliance decisions should still be of interest to accounting researchers. Though much of accounting research has focused on corporate settings, many policy questions for individuals and small businesses remain unanswered and are fertile grounds for exploration.

Finally, our measure of tax compliance focuses on the timing of tax payments and not about avoidance or evasion, which could include instances where individuals or corporation misreport their tax liability or simply not pay. Every measure of tax avoidance or compliance has their strengths and weaknesses. As noted in Hanlon and Heizman (2010), we urge readers to consider the differences in measures when addressing what can be learned from our study and generalized to other types of tax behavior.

References

2020 Public Services Trust 2010. IPSOS-MORI (2009), Voters not ready for spending cuts, https://www.ipsos-mori.com/researchpublications/researcharchive/poll.aspx (accessed April 2010).

Aaron, Henry, and Joel Slemrod, eds. *The crisis in tax administration*. Brookings Institution Press, 2004.

Allingham, Michael G., and Agnar Sandmo. "Income tax evasion: A theoretical analysis." *Journal of Public Economics* 1, no. 3-4 (1972): 323-338.

Alm, James, and Sarah Jacobson. "Using laboratory experiments in public economics." *National Tax Journal* (2007): 129-152.

Alm, James, and Benno Torgler. "Do ethics matter? Tax compliance and morality." *Journal of Business Ethics* 101, no. 4 (2011): 635-651.

Alm, James. "Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies." *International Tax and Public Finance* 19, no. 1 (2012): 54-77.

Alm, James, and Jorge Martinez-Vazquez. "Institutions, paradigms, and tax evasion in developing and transition countries." (2003).

Alós-Ferrer, Carlos, Alexander Jaudas, and Alexander Ritschel. "Effortful Bayesian updating: A pupil-dilation study." *Journal of Risk and Uncertainty* (2021): 1-22.

Andreoni, James, Brian Erard, and Jonathan Feinstein. "Tax compliance." *Journal of Economic Literature* 36, no. 2 (1998): 818-860.

Ariel, Barak. "Deterrence and moral persuasion effects on corporate tax compliance: findings from a randomized controlled trial." *Criminology* 50, no. 1 (2012): 27-69.

Atwood, T. J., Michael S. Drake, James N. Myers, and Linda A. Myers. "Home country tax system characteristics and corporate tax avoidance: International evidence."*The Accounting Review* 87, no. 6 (2012): 1831-1860.

Bach, Dominik R. "Skin conductance measures in neuroeconomic research." In *Neuroeconomics*, pp. 345-357. Springer, Berlin, Heidelberg, 2016.

Boame, Attah. "Individual Tax Compliance: A Time-Series Regression Using Canadian Data, 1987-2003." *Canada Revenue Agency Baseline Research Paper* (2008).

Boucsein, Wolfram. *Electrodermal activity*. Springer Science & Business Media, 2012.

Becker, Gary S. "Crime and punishment: An economic approach." In *The Economic Dimensions of Crime*, pp. 13-68. Palgrave Macmillan UK, (1968).

Blais, Ann-Renée, and Elke U. Weber. "A domain-specific risk-taking (DOSPERT) scale for adult populations." *Judgment and Decision Making* 1, no. 1 (2006): 33-47.

Bloomfield, Robert, Mark W. Nelson, and Eugene Soltes. "Gathering data for archival, field, survey, and experimental accounting research." *Journal of Accounting Research* 54, no. 2 (2016): 341-395.

Bloomquist, Kim M. "Tax evasion, income inequality and opportunity costs of compliance." In *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, vol. 96, pp. 91-104. National Tax Association, (2003).

Blumenthal, Marsha, Charles Christian, Joel Slemrod, and Matthew G. Smith. "Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota." *National Tax Journal* (2001): 125-138.

Bobek, Donna D., and Richard C. Hatfield. "An investigation of the theory of planned behavior and the role of moral obligation in tax compliance." *Behavioral Research in Accounting* 15, no. 1 (2003): 13-38.

Bradley, Margaret M., Laura Miccoli, Miguel A. Escrig, and Peter J. Lang. "The pupil as a measure of emotional arousal and autonomic activation." *Psychophysiology* 45, no. 4 (2008): 602-607.

Braithwaite, Valerie, and Eliza Ahmed. "A threat to tax morale: The case of Australian higher education policy." *Journal of Economic Psychology* 26, no. 4 (2005): 523-540.

Brink, Alisa G., Andrea Gouldman, Jacob M. Rose, and Kristian Rotaru. "Effects of superiors' compensation structures on psychophysiological responses and real earnings management decisions of subordinate managers." *Management Accounting Research* 48 (2020): 100691.

Cardinaels, Eddy, and Huaxiang Yin. "Think Twice Before Going for Incentives: Social Norms and the Principal's Decision on Compensation Contracts." *Journal of Accounting Research* 53, no. 5 (2015): 985-1015.

Cardinaels, Eddy, and Christoph Feichter. "Forced rating systems from employee and supervisor perspectives." *Journal of Accounting Research* 59, no. 5 (2021): 1573-1607.

Carillo, Paul E., Castro, Edgar, and Scartascini, Carlos. "Do rewards work? Evidence from the randomization of public works." IDB Working Paper IDB-WP-794 (2017).

Castro, Lucio, and Carlos Scartascini. "Tax compliance and enforcement in the pampas evidence from a field experiment." *Journal of Economic Behavior & Organization* 116 (2015): 65-82.

Chen, Kong-Pin, and CY Cyrus Chu. "Internal control versus external manipulation: A model of corporate income tax evasion." *Rand Journal of Economics* (2005): 151-164.

Chen, Yasheng, Johnny Jermias, and Tota Panggabean. "The role of visual attention in the managerial Judgment of Balanced ‐ Scorecard performance evaluation: Insights from using an eye ‐ tracking device." *Journal of Accounting Research* 54, no. 1 (2016): 113-146.

Christian, Roberta Calvet, and James Alm. "Empathy, sympathy, and tax compliance." *Journal of Economic Psychology* 40 (2014): 62-82.

Dalla Via, Nicola, Paolo Perego, and Marcel Van Rinsum. "How accountability type influences information search processes and decision quality." *Accounting, Organizations and Society* 75 (2019): 79-91.

De Neve, Jan-Emmanuel, Clement Imbert, Johannes Spinnewijn, Teodora Tsankova, and Maarten Luts. "How to improve tax compliance? Evidence from population-wide experiments in Belgium." *Journal of Political Economy* 129, no. 5 (2021): 1425-1463.

Deaton, Angus. "Saving and Liquidity Constraints."*Econometrica: Journal of the Econometric Society* (1991): 1221-1248.

Deci, Edward L. "Effects of externally mediated rewards on intrinsic motivation." *Journal of personality and Social Psychology* 18, no. 1 (1971): 105.

Doerrenberg, Philipp. "Does the use of tax revenue matter for tax compliance behavior?" *Economics Letters* 128 (2015): 30-34.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the European Economic Association* 9, no. 3 (2011): 522-550.

Elffers, Henk, Henry SJ Robben, and Dick J. Hessing. "On measuring tax evasion." *Journal of Economic Psychology* 13, no. 4 (1992): 545-567.

Erard, Brian, and Jonathan S. Feinstein. "Honesty and evasion in the tax compliance game." *The RAND Journal of Economics* (1994): 1-19.

Erard, Brian. "Taxation with representation: An analysis of the role of tax practitioners in tax compliance." *Journal of Public Economics* 52, no. 2 (1993): 163-197.

Feld, Lars P., Bruno S. Frey, and Benno Torgler. "Rewarding honest taxpayers? Evidence on the impact of rewards from field experiments." *Center for Research in Economics, Management and the Arts (CREMA) Working Paper* 16 (2006).

Feld, Lars P., and Claus Larsen. "Self-perceptions, government policies and tax compliance in Germany." *International Tax and Public Finance* 19, no. 1 (2012): 78-103.

Fehrenbacher, Dennis D., Axel K-D. Schulz, and Kristian Rotaru. "The moderating role of decision mode in subjective performance evaluation." *Management Accounting Research* 41 (2018): 1-10.

Floyd, Eric, and John A. List. "Using field experiments in accounting and finance." *Journal of Accounting Research* 54, no. 2 (2016): 437-475.

Frank, Björn, and Erich Kirchler. "Günter Schmölders and economic psychology: an introduction." In *The Psychology of Money and Public Finance*, pp. 1-16. Palgrave Macmillan UK, 2006.

Frey, B. S. “Not just for the money: An economic theory of personal motivation.” Cheltenham, UK: Edward Elgar (1997).

Gallemore, John, and Eva Labro. "The importance of the internal information environment for tax avoidance." *Journal of Accounting and Economics* 60, no. 1 (2015): 149-167.

Gollwitzer, Peter M., and Paschal Sheeran. "Implementation intentions and goal achievement: A meta-analysis of effects and processes." *Advances in Experimental Social Psychology* 38 (2006): 69-119.

Gollwitzer, Peter M. "Implementation intentions: Strong effects of simple plans." *American Psychologist* 54, no. 7 (1999): 493.

Gordon, James PP. "Individual morality and reputation costs as deterrents to tax evasion." *European Economic Review* 33, no. 4 (1989): 797-805.

Graham, John R., Michelle Hanlon, Terry Shevlin, and Nemit Shroff. "Incentives for tax planning and avoidance: Evidence from the field." *The Accounting Review* 89, no. 3 (2014): 991-1023.

Grasmick, Harold G., and Wilbur J. Scott. "Tax evasion and mechanisms of social control: A comparison with grand and petty theft. "*Journal of Economic Psychology* 2, no. 3 (1982): 213-230.

Gross, Tal, Matthew J. Notowidigdo, and Jialan Wang. "Online Appendix for “Liquidity Constraints and Consumer Bankruptcy: Evidence from Tax Rebates”." (2013).

Hallsworth, Michael. "The use of field experiments to increase tax compliance. "*Oxford Review of Economic Policy* 30.4 (2014): 658-679.

Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Vlaev. "The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. "*Journal of Public Economics* 148 (2017): 14-31.

Hanlon, Michelle, and Shane Heitzman. "A review of tax research. "*Journal of Accounting and Economics* 50, no. 2-3 (2010): 127-178.

Harrison, Glenn W., and John A. List. "Field experiments." *Journal of Economic literature* 42, no. 4 (2004): 1009-1055.

Hasan, Iftekhar, Chun-Keung Hoi, Qiang Wu, and Hao Zhang. "Does social capital matter in corporate decisions? Evidence from corporate tax avoidance." *Journal of Accounting Research* 55, no. 3 (2017): 629-668.

Hayes, Andrew F. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications, 2017.

Hayes, Taylor R., and Alexander A. Petrov. "Pupil diameter tracks the exploration–exploitation trade-off during analogical reasoning and explains individual differences in fluid intelligence." *Journal of Cognitive Neuroscience* 28, no. 2 (2016): 308-318.

Hayes, Andrew F., and Kristopher J. Preacher. "Statistical mediation analysis with a multicategorical independent variable." *British Journal of Mathematical and Statistical Psychology* 67, no. 3 (2014): 451-470.

Henderson, Robert R., Margaret M. Bradley, and Peter J. Lang. "Emotional imagery and pupil diameter." *Psychophysiology* 55, no. 6 (2018): e13050.

Her Majesty's Revenue and Customs (2010) Distraint: What it means for you. London, The Stationery Office.

Her Majesty's Revenue and Customs (2013) Levelling the playing field. London, The Stationery Office.

Hessing, Dick J., Henk Elffers, Henry SJ Robben, and Paul Webley. "Does deterrence deter? Measuring the effect of deterrence on tax compliance in field studies and experimental studies." *Why people pay taxes: Tax compliance and enforcement* (1992): 291-305.

Hoi, Chun Keung, Qiang Wu, and Hao Zhang. "Is corporate social responsibility (CSR) associated with tax avoidance? Evidence from irresponsible CSR activities." *The Accounting Review* 88, no. 6 (2013): 2025-2059.

Holz, Justin E., John A. List, Alejandro Zentner, Marvin Cardoza, and Joaquin Zentner. "The $100 million nudge: Increasing tax compliance of businesses and the self-employed using a natural field experiment". No. w27666. *National Bureau of Economic Research* (2020).

Hope, Ole-Kristian, Mark Shuai Ma, and Wayne B. Thomas. "Tax avoidance and geographic earnings disclosure." *Journal of Accounting and Economics* 56, no. 2-3 (2013): 170-189.

Howells, Fleur M., Dan J. Stein, and Vivienne A. Russell. "Perceived mental effort correlates with changes in tonic arousal during attentional tasks." *Behavioral and Brain Functions* 6, no. 1 (2010): 1-15.

Hyde, Jamiah, Katherine M. Ryan, and Allison M. Waters. "Psychophysiological markers of fear and anxiety." *Current Psychiatry Reports* 21, no. 7 (2019): 1-10.

IRS. 2006. Updated Estimates of the Tax Year 2001 Individual Income Tax Underreporting Gap. Overview. Internal Revenue Service: Washington DC.

Kagan, Robert A. "Understanding Regulatory Compliance." *LAW & Pol'y* 11 (1989): 89-103.

Komogortsev, Oleg V., Denise V. Gobert, Sampath Jayarathna, and Sandeep M. Gowda. "Standardization of automated analyses of oculomotor fixation and saccadic behaviors." *IEEE Transactions on Biomedical Engineering* 57, no. 11 (2010): 2635-2645.

Kirchler, Erich. *The economic psychology of tax behaviour*. Cambridge University Press, 2007.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. "Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark." *Econometrica* 79, no. 3 (2011): 651-69

Koo, Minjung, and Ayelet Fishbach. "Dynamics of self-regulation: How (un) accomplished goal actions affect motivation." *Journal of Personality and Social Psychology*, 94 (2008): 183–195.

Krishnamurthy, Parthasarathy, Patrick Carter, and Edward Blair. "Attribute framing and goal framing effects in health decisions." *Organizational Behavior and Human Decision Processes* 85, no. 2 (2001): 382-399.

Laeng, Bruno, and Dag Alnaes. "Pupillometry." In Klein, Christoph and Ettinger, Ulrich (Eds). *Eye Movement Research: An Introduction to Its Scientific Foundations and Applications*, pp. 449-502. Springer, Cham, 2019.

Laibson, David. "Golden eggs and hyperbolic discounting." *The Quarterly Journal of Economics* 112, no. 2 (1997): 443-478.

Levitt, Steven D., and John A. List. "What do laboratory experiments measuring social preferences reveal about the real world?" *The journal of Economic Perspectives* 21, no. 2 (2007): 153-174.

List, John A. "Social preferences: Some thoughts from the field." *Annu. Rev. Econ.* 1, no. 1 (2009): 563-579.

Lönnqvist, Jan-Erik, Markku Verkasalo, Gari Walkowitz, and Philipp C. Wichardt. "Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison." *Journal of Economic Behavior & Organization* 119 (2015): 254-266.

McGraw, Kathleen M., and John T. Scholz. "Appeals to civic virtue versus attention to self-interest: Effects on tax compliance." *Law and Society Review* (1991): 471-498.

McGuire, Sean T., Thomas C. Omer, and Dechun Wang. "Tax avoidance: Does tax-specific industry expertise make a difference?" *The Accounting Review* 87, no. 3 (2012): 975-1003.

Murphy, Kristina. "Regulating More Effectively: The relationship between procedural justice, legitimacy, and tax non-compliance." *Journal of Law and Society* 32, no. 4 (2005): 562-589.

Nunes, Joseph C., and Xavier Drèze. "The endowed progress effect: How artificial advancement increases effort." *Journal of Consumer Research* 32, no. 4 (2006): 504-512.

O'Donoghue, Ted, and Matthew Rabin. "Doing it now or later." *American Economic Review* (1999): 103-124.

OECD. "Understanding and influencing Taxpayers' Compliance Behavior." Paris, France (2010).

O'Keefe, Daniel J., and Jakob D. Jensen. "The relative effectiveness of gain-framed and loss-framed persuasive appeals concerning obesity-related behaviors: Meta-analytic evidence and implications." *Leveraging consumer psychology for effective health communications: The obesity challenge* (2011): 171-185.

Ortega, Daniel & Sanguinetti, Pablo. "Deterrence and reciprocity effects on tax compliance: Experimental evidence from Venezuela." *Working Paper number 2013/08*. Caracas, CAF Development Bank (2013).

Orviska, Marta, and John Hudson. "Tax evasion, civic duty and the law abiding citizen." *European Journal of Political Economy* 19, no. 1 (2003): 83-102.

Phelps, Edmund S., and Robert A. Pollak. "On second-best national saving and game-equilibrium growth." *The Review of Economic Studies* 35, no. 2 (1968): 185-199.

Pomeranz, Dina. "No taxation without information: Deterrence and self-enforcement in the value added tax." *The American Economic Review* 105, no. 8 (2015): 2539-2569.

Reckers, Philip MJ, Debra L. Sanders, and Stephen J. Roark. "The influence of ethical attitudes on taxpayer compliance." *National Tax Journal* (1994): 825-836.

Rose, Anna M., Jacob M. Rose, Kristian Rotaru, Kerri-Ann Sanderson, and Jay Thibodeau. "Effects of uncertainty visualization on attention arousal and Judgment." *Behavioral Research in Accounting* (2022, forthcoming).

Rose, Anna M., Jacob M. Rose, Kristian Rotaru, Kerri-Ann Sanderson, and Jay Thibodeau. "Effects of data visualization choices on psychophysiological responses, auditor judgment and audit quality." *The Journal of Information Systems* (2021, forthcoming).

Rotaru, Kristian, Kalev, Petko, Yadav, Nitin, and Peter Bossaerts. "Transferring Cognitive Talent Across Domains to Reduce the Disposition Effect in Investment." *Scientific Reports* 11, 23068 (2021).

Roth, Jeffrey A., and John T. Scholz. *Taxpayer Compliance, Volume 1: An Agenda for Research*. Vol. 1. University of Pennsylvania Press, 1989.

Sandmo, Agnar. "The theory of tax evasion: A retrospective view." *National Tax Journal* (2005): 643-663.

Schmölders, Günter. "Fiscal psychology: A new branch of public finance." *National Tax Journal* 12, no. 4 (1959): 340-345.

Sirois, Sylvain, and Julie Brisson. "Pupillometry." *Wiley Interdisciplinary Reviews: Cognitive Science* 5, no. 6 (2014): 679-692.

Slemrod, Joel, and Caroline Weber. "Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy." *International Tax and Public Finance* 19, no. 1 (2012): 25-53.

Smith, K. W. "Reciprocity and Fairness: Positive Incentives for Tax Compliance Why People Pay Taxes." (1992).

Srinivasan, Thirukodikaval N. "Tax evasion: A model." *Journal of Public Economics* 2, no. 4 (1973): 339-346.

Tank, Ann K., and Anne M. Farrell. "Is Neuroaccounting Taking a Place on the Stage? A Review of the Influence of Neuroscience on Accounting Research." *European Accounting Review* (2021): 1-35.

Tayler, William B., and Robert J. Bloomfield. "Norms, conformity, and controls." *Journal of Accounting Research* 49, no. 3 (2011): 753-790.

Taylor, Shelley E. "Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis." *Psychological Bulletin* 110, no. 1 (1991): 67.

Tobii (2014). Tobii TX300 Eye Tracker User Manual, Tobii Technology.

Trivedi, V. Umashanker, Mohamed Shehata, and Bernadette Lynn. "Impact of personal and situational factors on taxpayer compliance: An experimental analysis." *Journal of Business Ethics* 47, no. 3 (2003): 175-197.

Trivedi, V.U., Shehata, M. & Mestelman, S. "Attitudes, Incentives, and Tax Compliance." *Canadian Tax Journal* 53, (2005) 29-61.

Unsworth, Nash, and Matthew K. Robison. "The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43, no. 12 (2017): 1962.

Wang, Hung-Jen, and Michelle J. White. "An optimal personal bankruptcy procedure and proposed reforms." *The Journal of Legal Studies* 29, no. 1 (2000): 255-286.

Wang, Joseph Tao-yi, Michael Spezio, and Colin F. Camerer. "Pinocchio's pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games." *American Economic Review* 100, no. 3 (2010): 984-1007.

Webley, Paul. *Tax evasion: An experimental approach*. Cambridge University Press, 1991.

Wenzel, Michael. "Misperceptions of social norms about tax compliance: From theory to intervention." *Journal of Economic Psychology* 26, no. 6 (2005): 862-883.

Wenzel, Michael. "A letter from the tax office: Compliance effects of informational and interpersonal justice." *Social Justice Research* 19, no. 3 (2006): 345-364.

Wiebenga, Jacob H., and Bob M. Fennis. "The road traveled, the road ahead, or simply on the road? When progress framing affects motivation in goal pursuit☆." *Journal of Consumer Psychology* 24, no. 1 (2014): 49-62.

Yitzhaki, Shlomo. "Income tax evasion: A theoretical analysis." *Journal of Public Economics* 3, no. 2 (1974): 201-202.

Table 1: Background characteristics - Experiment I

Group name	Test phrase	N	Debt value	Mean debt value	Mean Age	% Male
Control		8,910	£27,642,738.92	£3,102.44	49.88	72.85
Reminder	Have you overlooked this payment? In case you might have done, I'm writing to give you a reminder.	8,835	£29,033,619.62	£3,286.20	49.47	72.61
Moral Duty	Paying your tax is the right thing to do.	8,677	£27,509,599.64	£3,170.40	49.51	73.04
Moral Equity	Paying your tax is the fair thing to do.	8,497	£26,667,017.63	£3,138.40	49.87	72.20
Public Services	According to a 2009 opinion poll, our most valued public services are the NHS, schools, care for the elderly, and the police. Every single tax payment helps to run these services.	8,500	£25,738,916.22	£3,028.11	49.45	72.34
Costs & Gain	Every tax payment we receive means the country has to borrow less money – reducing costs for us all.	8,714	£28,510,852.97	£3,271.84	49.52	72.70
Costs & Loss	Every tax payment we do not receive means the country has to borrow more money - increasing costs for us all.	8,632	£27,272,137.56	£3,159.42	49.83	72.24
Deficit	According to a 2011 opinion survey, people said that one of the best ways to improve the economy was to reduce the public debt and deficit. But to reduce the deficit we need everyone to pay the tax they owe. Please help us achieve this goal by making your payment now.	8,145	£25,814,753.72	£3,169.40	49.93	71.82
Progress	You have already successfully filed your tax return. All you need to do now is call us to pay the amount you said you owe.	8,865	£28,627,901.69	£3,229.32	49.78	72.05
Progress & Thanks	Thank you for successfully filing your 2010-11 tax return: we appreciate this requires effort on your part. All you need to do now is call us to pay the amount you said you owe.	7,984	£24,912,405.24	£3,120.29	50.01	72.43
Monitoring	We will be checking how long it takes you to respond to this letter.	8,734	£27,369,606.23	£3,133.69	50.08	73.72
Monitoring Day	We will be checking our records every day to see if you have paid.	8,695	£27,573,397.84	£3,171.18	49.88	71.77
Monitoring & Norms	Nine out of ten people pay their tax on time - you are currently in the small minority of people that have not paid us yet. We will be checking how long it takes you to respond to this letter.	7,855	£25,339,186.83	£3,225.87	49.83	71.97
Moral & Monitoring & Norms	Paying your tax is the right thing to do, and nine out of ten people pay their tax on time. You are currently in the small minority of people that have not paid us yet. We will be checking how long it takes you to respond to this letter.	8,685	£27,201,634.73	£3,132.02	50.68	72.54

Table 2: Date of letter issue by group - Experiment I

Group name	Day of issue						Total issue
	Mon Feb 27th	Tue Feb 28th	Wed Feb 29th	Thu Mar 1st	Fri Mar 2nd	Mon Mar 6th	
Control	1,730	1,409	1,678	1,394	1,281	1,410	9,896
Reminder	1,633	1,773	1,377	1,463	1,312	1,269	9,822
Moral Duty	1,664	1,573	1,383	1,371	1,307	1,375	9,672
Moral Equity	1,751	1,306	1,566	1,128	1,422	1,319	9,490
Public Services	1,534	1,463	1,650	1,461	1,413	1,189	9,713
Costs & Gain	1,671	1,487	1,466	1,370	1,270	1,233	9,497
Costs & Loss	1,484	1,922	1,453	1,280	1,276	1,315	9,731
Deficit	1,720	1,559	1,297	1,451	1,373	1,283	9,685
Progress	1,476	1,742	1,451	1,449	1,528	1,214	9,856
Progress & Thanks	1,712	1,781	1,422	569	1,424	1,044	8,949
Monitoring	1,652	1,684	1,490	1,140	1,278	1,382	9,630
Monitoring Day	1,597	1,327	1,508	1,233	1,401	1,076	9,147
Monitoring & Norms	1,576	1,580	695	1,434	1,409	1,154	8,854
Moral & Monitoring & Norms	1,717	1,535	1,424	1,459	1,279	1,266	9,687

Notes: In this table the number of "Progress & Thanks" letters listed as issued on 1st March, and "Monitoring & Norms" letters issued on 29th February, are clearly outliers. This is because of isolated technical difficulties relating to the program used to identify cases.

Figure 1: Raw data on the cumulative percentage of people paying per day by treatment group for the first 23 days, Experiment I

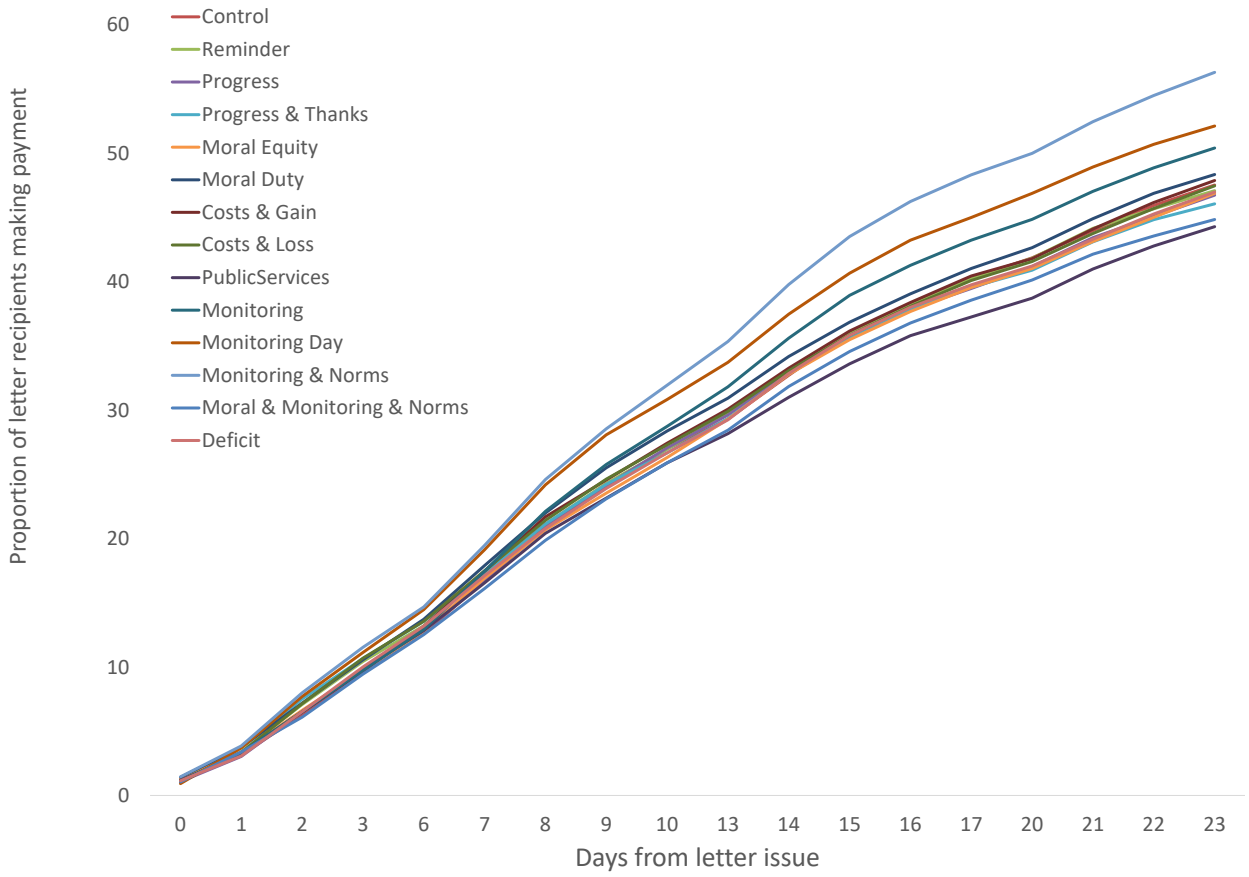


Table 3: Logistic regression on tax payments - Experiment I

	(I) Pay tax	(II) Pay tax	(III) Pay tax – recent debtors	(IV) Pay tax – no recent debt
Reminder	0.005	0.006	-0.005	0.026*
	(0.008)	(0.008)	(0.010)	(0.013)
Moral Duty	0.001	0.002	-0.016	0.031*
	(0.008)	(0.008)	(0.010)	(0.013)
Moral Equity	0.007	0.007	0.003	0.017
	(0.008)	(0.008)	(0.010)	(0.013)
Public Services	0.002	0.001	0.007	-0.004
	(0.008)	(0.008)	(0.010)	(0.014)
Costs & Gain	-0.009	-0.008	-0.012	-0.000
	(0.008)	(0.008)	(0.010)	(0.013)
Costs & Loss	0.004	0.003	0.001	0.006
	(0.008)	(0.008)	(0.010)	(0.013)
Deficit	-0.007	-0.007	-0.004	-0.008
	(0.008)	(0.008)	(0.010)	(0.013)
Progress	-0.012	-0.012	-0.013	-0.008
	(0.008)	(0.008)	(0.010)	(0.013)
Progress &	-0.010	-0.011	-0.015	0.001
Thanks	(0.008)	(0.008)	(0.010)	(0.014)
Monitoring	0.049***	0.051***	0.048***	0.055***
	(0.008)	(0.008)	(0.010)	(0.013)
Monitoring Day	0.026***	0.025***	0.018	0.041**
	(0.008)	(0.008)	(0.010)	(0.013)
Monitoring &	0.044***	0.044***	0.044***	0.045***
Norms	(0.008)	(0.008)	(0.010)	(0.013)
Moral &	0.037***	0.037***	0.033	0.044**
Monitoring &	(0.008)	(0.008)	(0.010)	(0.013)
Norms				
Age		0.003***	0.002***	0.003***
		(0.000)	(0.000)	(0.000)
Male		-0.052***	-0.039***	-0.055***
		(0.003)	(0.004)	(0.005)
Debt size		-5.15e-07	-3.41e-07	-1.08e-06*
		(0.000)	(0.000)	(0.000)
Self-employed		0.020***	0.022**	0.020*
		(0.006)	(0.007)	(0.009)
Accountant		-0.031***	-0.025***	-0.015**
		(0.003)	(0.005)	(0.006)
Constant		0.57***	0.54***	0.62***
		(0.008)	(0.010)	(0.013)
N	105,379	105,000	68,274	36,726
Pseudo R^2	0.00	0.01	0.00	0.00

* p<0.05, ** p<0.01, *** p<0.001

Notes: Our dependent variable is whether a recipient started to pay or paid in full their outstanding tax within the 19-day period. Standard errors are presented in parentheses. The sample sizes are different in I vs II, and III vs IV because the full range of covariates were not present for every record. “Recent debtors” refer to those individuals who have incurred a debt within any of the preceding three tax years.

Table 4: Logistic regression on tax payments in Experiment I, by background characteristics

	(I) Male	(II) Female	(III) Below median age	(IV) Above median age	(V) Accountant	(VI) No accountant
Reminder	0.010	0.002	0.011	0.002	0.033	0.015
	(0.009)	(0.016)	(0.012)	(0.011)	(0.009)	(0.018)
Moral Duty	0.009	-0.016	0.006	-0.003	-0.015	0.009
	(0.009)	(0.016)	(0.012)	(0.011)	(0.009)	(0.018)
Moral Equity	0.011	-0.005	0.013	0.004	0.004	0.025
	(0.009)	(0.016)	(0.012)	(0.011)	(0.009)	(0.018)
Public	0.001	0.008	0.011	-0.006	-0.002	0.021
Services	(0.010)	(0.016)	(0.012)	(0.011)	(0.009)	(0.018)
Costs & Gain	-0.004	-0.022	-0.015	-0.002	-0.010	-0.006
	(0.009)	(0.016)	(0.012)	(0.011)	(0.009)	(0.018)
Costs & Loss	0.007	-0.010	-0.004	0.012	0.006	-0.006
	(0.009)	(0.016)	(0.012)	(0.011)	(0.009)	(0.018)
Deficit	0.003	-0.028	-0.002	-0.010	-0.010	0.005
	(0.009)	(0.016)	(0.012)	(0.011)	(0.009)	(0.018)
Progress	-0.006	-0.022	-0.013	-0.009	-0.017*	0.014
	(0.009)	(0.016)	(0.012)	(0.011)	(0.009)	(0.018)
Progress &	-0.004	-0.022	-0.010	-0.009	-0.015	-0.001
Thanks	(0.010)	(0.016)	(0.012)	(0.011)	(0.009)	(0.019)
Monitoring	0.058***	0.026	0.050***	0.051***	0.050***	0.046
	(0.009)	(0.015)	(0.012)	(0.011)	(0.009)	(0.018)
Monitoring	0.039***	-0.005	0.032**	0.021*	0.025**	0.028
Day	(0.009)	(0.016)	(0.012)	(0.011)	(0.009)	(0.018)
Monitoring &	0.051***	0.029	0.046***	0.043***	0.046***	0.031
Norms	(0.009)	(0.015)	(0.012)	(0.011)	(0.009)	(0.018)
Moral &	0.052***	-0.001	0.038**	0.037***	0.032***	0.062**
Monitoring &	(0.009)	(0.016)	(0.012)	(0.011)	(0.009)	(0.018)
Norms						
Constant	0.555	0.604	0.546	0.561	0.565	0.594
N	76,094	26,869	49,195	56,184	86,075	19,304
Pseudo R^2	0.00	0.00	0.00	0.00	0.00	0.00

* p<0.05, ** p<0.01, *** p<0.001

Notes: Our dependent variable is whether a recipient started to pay or paid in full their outstanding tax within the 19-day period. Standard errors are presented in parentheses.

Table 5: Logistic regression on tax payments in Experiment I, by debt quartiles

	(I) First Debt Quartile	(II) Second Debt Quartile	(III) Third Debt Quartile	(IV) Fourth Debt Quartile
Reminder	0.005	0.002	0.011	0.010
	(0.015)	(0.016)	(0.012)	(0.011)
Moral Duty	0.014	-0.016	0.006	-0.003
	(0.015)	(0.016)	(0.012)	(0.011)
Moral Equity	-0.014	-0.005	0.013	0.004
	(0.016)	(0.016)	(0.012)	(0.011)
Public	0.011	0.008	0.011	-0.006
Services	(0.016)	(0.016)	(0.012)	(0.011)
Costs & Gain	0.008	-0.022	-0.015	-0.002
	(0.015)	(0.016)	(0.012)	(0.011)
Costs & Loss	0.011	-0.010	-0.004	0.012
	(0.015)	(0.016)	(0.012)	(0.011)
Deficit	-0.006	-0.028	-0.002	-0.010
	(0.015)	(0.016)	(0.012)	(0.011)
Progress	0.010	-0.022	-0.013	-0.009
	(0.015)	(0.016)	(0.012)	(0.011)
Progress &	-0.009	-0.022	-0.010	-0.009
Thanks	(0.016)	(0.016)	(0.012)	(0.011)
Monitoring	0.057***	0.026	0.050***	0.051***
	(0.015)	(0.015)	(0.012)	(0.011)
Monitoring	0.025	-0.005	0.032**	0.021*
Day	(0.009)	(0.016)	(0.012)	(0.011)
Monitoring &	0.044**	0.029	0.046***	0.043***
Norms	(0.015)	(0.015)	(0.012)	(0.011)
Moral &	0.047**	-0.001	0.038**	0.037***
Monitoring &	(0.015)	(0.016)	(0.012)	(0.011)
Norms				
Constant	0.598	0.604	0.546	0.560
N	27,431	26,869	49,195	25,513
Pseudo R^2	0.00	0.00	0.00	0.00

* p<0.05, ** p<0.01, *** p<0.001

Notes: Our dependent variable is whether a recipient started to pay or paid in full their outstanding tax within the 19-day period. Standard errors are presented in parentheses.

Table 6: Logistic regression on paying tax within 42 and 70 days, Experiment I

	(I) Pay tax in 42 days	(II) Pay tax in 70 days
Reminder	0.011	0.003
	(0.007)	(0.005)
Moral Duty	0.013*	0.013*
	(0.007)	(0.005)
Moral Equity	0.010	0.004
	(0.007)	(0.005)
Public Services	0.007	0.003
	(0.007)	(0.005)
Costs & Gain	0.008	0.009
	(0.007)	(0.005)
Costs & Loss	0.011	0.014*
	(0.007)	(0.005)
Deficit	0.013*	0.012*
	(0.007)	(0.005)
Progress	0.001	0.009
	(0.007)	(0.005)
Progress & Thanks	0.004	0.010
	(0.007)	(0.005)
Monitoring	0.033***	0.018***
	(0.006)	(0.005)
Monitoring Day	0.014*	0.005
	(0.007)	(0.005)
Monitoring & Norms	0.031***	0.016**
	(0.006)	(0.005)
Moral & Monitoring & Norms	0.018**	0.010
	(0.007)	(0.005)
Age	0.002***	0.002***
	(0.000)	(0.000)
Male	-0.030***	-0.014***
	(0.003)	(0.002)
Debt size	-3.72e-06***	-2.53e-06***
	(0.000)	(0.000)
Self-employed	0.025***	0.018***
	(0.005)	(0.004)
Accountant	-0.005	0.003***
	(0.003)	(0.003)
Constant	0.727***	0.830***
Pseudo R2	0.05	0.05
N	116,148	119,303

* p<0.05, ** p<0.01, *** p<0.001

Notes: Our dependent variable is whether a recipient started to pay or paid in full their outstanding tax. Standard errors are presented in parentheses.

Table 7: Background characteristics - Experiment II

Group name	Test phrase	N	Debt value	Mean debt value	Mean Age	% Male
Control		17,208	£43,066,664.71	£2,503.29	49.57	71.34
Interest Today	We are charging you interest daily. You will save money if you pay now.	17,061	£42,772,379.82	£2,507.61	49.58	72.00
Interest Tomorrow	We are charging you interest daily. You will lose money if you pay later.	16,866	£41,703,233.28	£2,473.35	49.58	70.70
Less Effort	Paying now means less hassle because we won't contact you about this again.	17,172	£43,911,399.60	£2,557.75	48.99	71.44
More Effort	Paying later means more hassle because we will contact you about this again.	16,960	£42,966,785.52	£2,533.87	49.36	71.63
Costs Further Action	If you pay this debt now, you can stop us taking any further action.	17,048	£43,138,652.55	£2,531.02	49.26	70.46
Costs Peace	If you pay this debt now, you won't have to think about it anymore.	17,162	£42,752,837.97	£2,491.57	49.35	71.92
Interest & Effort	We are charging you interest daily. You will save money if you pay today. Paying now means less hassle because we won't contact you about this again.	17,134	£42,621,810.33	£2,487.99	49.45	71.96
Interest & Action	We are charging you interest daily. You will save money if you pay today. If you pay this debt now, you can stop us taking any further action.	16,858	£42,738,043.43	£2,535.63	49.18	71.58
Contact Now	Please contact us as soon as you receive this letter.	16,960	£42,323,999.15	£2,491.55	49.20	71.63
Contact Now Information	Please contact us as soon as you receive this letter. We are open Monday to Saturday 8am to 8pm, Sundays 8am to 5pm - or you can pay online any time.	17,124	£43,657,948.21	£2,548.48	49.25	71.54
Contact Later Information	Please choose a time tomorrow to contact us. We are open Monday to Saturday 8am to 8pm, Sundays 8am to 5pm - or you can pay online any time.	17,348	£43,663,133.99	£2,517.33	49.29	71.66

Table 8: Date of letter issue by group - Experiment II

Group name	Day of issue							Total issue
	Tue Mar 5th 2013	Wed Mar 6th 2013	Thu Mar 7th 2013	Fri Mar 8th 2013	Mon Mar 11th 2013	Tue Mar 12th 2013	Wed Mar 13th 2013	
Control	2,913	2,808	2,651	2,579	2,575	2,482	1,200	17,208
Interest Today	2,985	2,784	2,646	2,584	2,483	2,420	1,159	17,061
Interest Tomorrow	2,947	2,704	2,679	2,489	2,379	2,441	1,227	16,866
Less Effort	3,070	2,809	2,697	2,531	2,436	2,469	1,160	17,172
More Effort	3,012	2,741	2,609	2,568	2,431	2,401	1,198	16,960
Costs Further Action	2,971	2,852	2,656	2,516	2,493	2,465	1,095	17,048
Costs Peace	3,033	2,711	2,684	2,599	2,492	2,462	1,181	17,162
Interest & Effort	3,072	2,767	2,679	2,641	2,392	2,413	1,170	17,134
Interest & Action	2,924	2,775	2,607	2,552	2,370	2,481	1,149	16,858
Contact Now	2,957	2,755	2,659	2,583	2,431	2,440	1,169	16,994
Contact Now Information	2,978	2,740	2,735	2,533	2,452	2,453	1,243	17,134
Contact Later Information	3,008	2,727	2,718	2,669	2,485	2,455	1,286	17,348

Table 9: Logistic regression on paying tax - Experiment II

	(I) Pay tax – Full sample	(II) Pay tax – Full sample	(III) Pay tax – recent debtors	(IV) Pay tax – no recent debt
Interest Today	0.000	0.001	-0.013*	0.025**
	(0.005)	(0.005)	(0.007)	(0.009)
Interest Tomorrow	0.008	0.008	-0.002	0.024**
	(0.005)	(0.005)	(0.007)	(0.009)
Less Effort	-0.003	-0.01	-0.005	0.001
	(0.005)	(0.005)	(0.007)	(0.009)
More Effort	0.002	0.003	-0.004	0.012
	(0.005)	(0.005)	(0.007)	(0.009)
Costs Further Action	0.044***	0.046***	0.038***	0.057***
	(0.005)	(0.006)	(0.007)	(0.009)
Costs Peace	-0.008	-0.007	-0.009	-0.006
	(0.005)	(0.005)	(0.007)	(0.009)
Interest & Effort	0.006	0.009	-0.003	0.022*
	(0.005)	(0.005)	(0.007)	(0.009)
Interest & Action	0.017**	0.019***	0.014*	0.024**
	(0.005)	(0.005)	(0.007)	(0.009)
Contact Now	0.022***	0.024***	0.022***	0.021*
	(0.005)	(0.005)	(0.007)	(0.009)
Contact Now Information	0.008	0.010	0.013*	0.021*
	(0.005)	(0.005)	(0.007)	(0.009)
Contact Later Information	0.029***	0.031***	0.030***	0.027**
	(0.005)	(0.005)	(0.007)	(0.009)
Age		0.004***		
		(0.000)		
Male		-0.032***		
		(0.003)		
Debt size		-4.59e-06***		
		(0.000)		
Self-employed		0.121***		
		(0.004)		
Accountant		-0.004		
		(0.003)		
Experienced		-0.048***		
		(0.002)		
Constant	0.421***	0.420***	0.405***	0.452***
	(0.016)	(0.031)	(0.019)	(0.026)
Pseudo R2	0.00	0.01	0.00	0.00
N	204,936	204,367	133,665	71,271

* p<0.05, ** p<0.01, *** p<0.001.

Notes: Our dependent variable for I - IV is whether the taxpayer started to pay or paid in full their outstanding tax within the 23-day period. Standard errors are presented in parentheses. The sample sizes are different in I vs II because not everyone has data on age or gender.

Figure 2: Raw data on the cumulative percentage of people paying per day by treatment group for the first 48 days, Experiment II

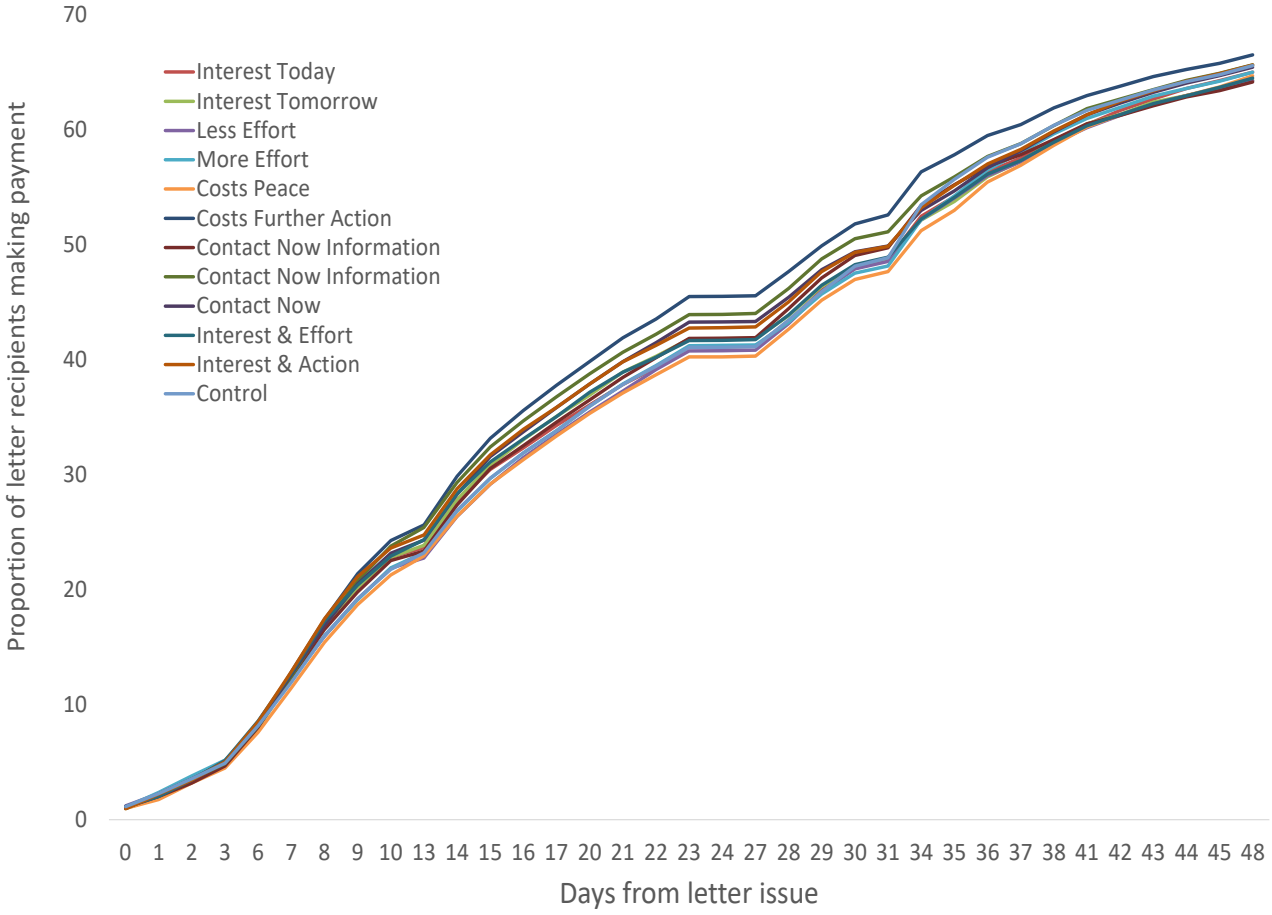


Table 10: Logistic regression on tax payments in Experiment II, by background characteristics

	(I) Male	(II) Female	(III) Below median age	(IV) Above median age	(V) Accountant	(VI) No accountant
Interest Today	-0.007	0.015	-0.017*	0.018*	-0.008	0.025*
	(0.006)	(0.011)	(0.008)	(0.008)	(0.006)	(0.011)
Interest Tomorrow	0.004	0.014	0.001	0015*	0.007	0.008
	(0.006)	(0.010)	(0.008)	(0.008)	(0.006)	(0.011)
Less Effort	-0.008	0.010	-0.009	0.004	-0.002	-0.007
	(0.006)	(0.010)	(0.008)	(0.008)	(0.006)	(0.011)
More Effort	-0.002	0.009	-0.009	0.012	0.002	0.001
	(0.006)	(0.011)	(0.008)	(0.008)	(0.006)	(0.011)
Costs Further Action	0.042***	0.049***	0.039***	0.051***	0.046***	0.039***
	(0.006)	(0.010)	(0.008)	(0.008)	(0.008)	(0.011)
Costs Peace	-0.012*	-0.001	-0.021**	0.005	-0.008	-0.009
	(0.006)	(0.011)	(0.008)	(0.008)	(0.006)	(0.011)
Interest & Effort	-0.003	0.021	0.006	0.006	0.003	0.013
	(0.006)	(0.011)	(0.008)	(0.008)	(0.006)	(0.011)
Interest & Action	0.012	0.029**	0.011	0.022**	0.016**	0.018
	(0.006)	(0.011)	(0.008)	(0.008)	(0.006)	(0.011)
Contact Now	0.019**	0.034**	0.016*	0.030***	0.023***	0.019
	(0.006)	(0.011)	(0.008)	(0.008)	(0.006)	(0.011)
Contact Now Information	0.004	0.016	0.007	0.009	0.015*	-0.012
	(0.006)	(0.011)	(0.008)	(0.008)	(0.008)	(0.011)
Contact Later Information	0.026***	0.034**	0.019*	0.039***	0.032***	0.020
	(0.006)	(0.010)	(0.008)	(0.008)	(0.006)	(0.011)
Constant	0.410***	0.446***	0.456***	0.386***	0.420***	0.424***
N	146,484	54,320	102,686	102,250	153,007	51,929
Pseudo R^2	0.00	0.00	0.00	0.00	0.00	0.00

* p<0.05, ** p<0.01, *** p<0.001.

Notes: Our dependent variable is whether a recipient started to pay or paid in full their outstanding tax within the initial 23-day period. Standard errors are presented in parentheses.

Table 11: Logistic regression on tax payments in Experiment II, by debt quartiles

	(I) First Debt Quartile	(II) Second Debt Quartile	(III) Third Debt Quartile	(IV) Fourth Debt Quartile
Interest Today	0.003	0.014	-0.010	-0.006
	(0.011)	(0.011)	(0.011)	(0.010)
Interest Tomorrow	0.016	0.016	0.003	-0.006
	(0.011)	(0.011)	(0.011)	(0.010)
Less Effort	-0.008	0.002	0.007	-0.015
	(0.011)	(0.011)	(0.011)	(0.010)
More Effort	-0.001	0.006	-0.001	0.001
	(0.011)	(0.011)	(0.011)	(0.011)
Costs Further Action	0.049***	0.064***	0.035***	0.029**
	(0.011)	(0.011)	(0.011)	(0.011)
Costs Peace	-0.004	0.003	-0.016	-0.017
	(0.011)	(0.011)	(0.011)	(0.010)
Interest & Effort	0.023*	0.020	-0.008	-0.013
	(0.011)	(0.011)	(0.011)	(0.010)
Interest & Action	0.036**	0.021	0.012	-0.004
	(0.011)	(0.011)	(0.011)	(0.010)
Contact Now	0.037**	0.019	0.024*	0.009
	(0.011)	(0.011)	(0.011)	(0.011)
Contact Now Information	0.006	0.014	0.003	0.007
	(0.011)	(0.011)	(0.011)	(0.011)
Contact Later Information	0.028**	0.035**	0.021*	0.030**
	(0.011)	(0.011)	(0.011)	(0.011)
Constant	0.460	0.422	0.420	0.380
N	51,224	51,223	51,222	51,268
Pseudo R^2	0.00	0.00	0.00	0.00

* p<0.05, ** p<0.01, *** p<0.001.

Notes: Our dependent variable is whether a recipient started to pay or paid in full their outstanding tax within the initial 23-day period. Standard errors are presented in parentheses.

Table 12: Logistic regression on paying tax within 48 and 70 days, Experiment II

	(I) Pay tax in 48 days	(II) Pay tax in 70 days
Interest Today	-0.006	-0.07
	(0.005)	(0.005)
Interest Tomorrow	-0.013*	-0.013*
	(0.005)	(0.005)
Less Effort	-0.013*	-0.011*
	(0.005)	(0.005)
More Effort	-0.005	-0.002
	(0.005)	(0.005)
Costs Further Action	0.010	0.004
	(0.005)	(0.005)
Costs Peace	-0.009	-0.006
	(0.005)	(0.005)
Interest & Effort	-0.009	-0.006
	(0.005)	(0.005)
Interest & Action	0.001	0.001
	(0.005)	(0.005)
Contact Now	-0.001	-0.001
	(0.005)	(0.005)
Contact Now Information	-0.013*	-0.019***
	(0.005)	(0.005)
Contact Later Information	0.001	0.001
	(0.005)	(0.005)
Age	0.003***	0.002***
	(0.000)	(0.000)
Male	-0.022***	-0.019***
	(0.002)	(0.002)
Debt size	-1.36e-06***	-3.42e-07***
	(0.000)	(0.000)
Self-employed	0.146***	0.123***
	(0.004)	(0.004)
Accountant	0.005	0.003
	(0.002)	(0.002)
Experienced	0.001	0.011***
	(0.002)	(0.002)
Constant	0.651	0.710
Pseudo R2	0.00	0.00
N	204,367	204,367

* p<0.05, ** p<0.01, *** p<0.001

Notes: Our dependent variable is whether a recipient started to pay or paid in full their outstanding tax. Standard errors are presented in parentheses.

Table 13: Background characteristics - Experiment III

Group name	Test phrase	N	Likelihood of repayment	Punishment severity	Realism	Age	% Female	Risk propensity
Costs Further Action	We told you recently that you have an outstanding tax obligation with the government. If you pay this debt now, you can stop us taking any further action.	1,010	8.53 (1.93)	6.52 (1.92)	6.88 (2.19)	36.75 (11.54)	60.69	4.50 (2.38)
Fine Non-specific	We told you recently that you have an outstanding tax obligation with the government. If you do not pay your taxes, you may be subject to a fine.	1,005	8.59 (1.93)	6.41 (1.91)	7.60 (2.11)	36.27 (11.05)	62.29	4.53 (2.31)
Jailtime Non-specific	We told you recently that you have an outstanding tax obligation with the government. If you do not pay your taxes, you may be subject to prosecution that results in time in prison.	1,010	8.73 (1.77)	6.99 (1.95)	6.37 (2.44)	36.45 (10.84)	61.68	4.49 (2.36)
Fine Average	We told you recently that you have an outstanding tax obligation with the government. If you do not pay your taxes, we could fine you 5% of your current outstanding tax obligation.	1,008	8.51 (1.91)	6.10 (2.25)	7.02 (2.33)	37.44 (11.14)	58.53	4.44 (2.28)
Jailtime Average	We told you recently that you have an outstanding tax obligation with the government. If you do not pay your taxes, you could face 6 months in prison.	1,014	8.77 (1.80)	6.91 (2.16)	5.83 (2.69)	36.36 (11.01)	60.75	4.57 (2.36)
Fine Maximum	We told you recently that you have an outstanding tax obligation with the government. If you do not pay your taxes, you could face a fine of up to 100% of the outstanding tax obligation and the possibility that the government seizes your personal assets (e.g. property).	1,007	8.75 (1.71)	7.72 (1.76)	7.00 (2.38)	37.13 (11.11)	58.09	4.43 (2.31)
Jailtime Maximum	We told you recently that you have an outstanding tax obligation with the government. If you do not pay your taxes, you could face up to 7 years in prison.	1,010	8.66 (1.78)	6.89 (2.30)	5.60 (2.76)	36.89 (11.21)	57.72	4.43 (2.38)

Notes: Standard deviations are presented in parentheses.

Figure 3: Repayment likelihood by treatment group – Experiment III

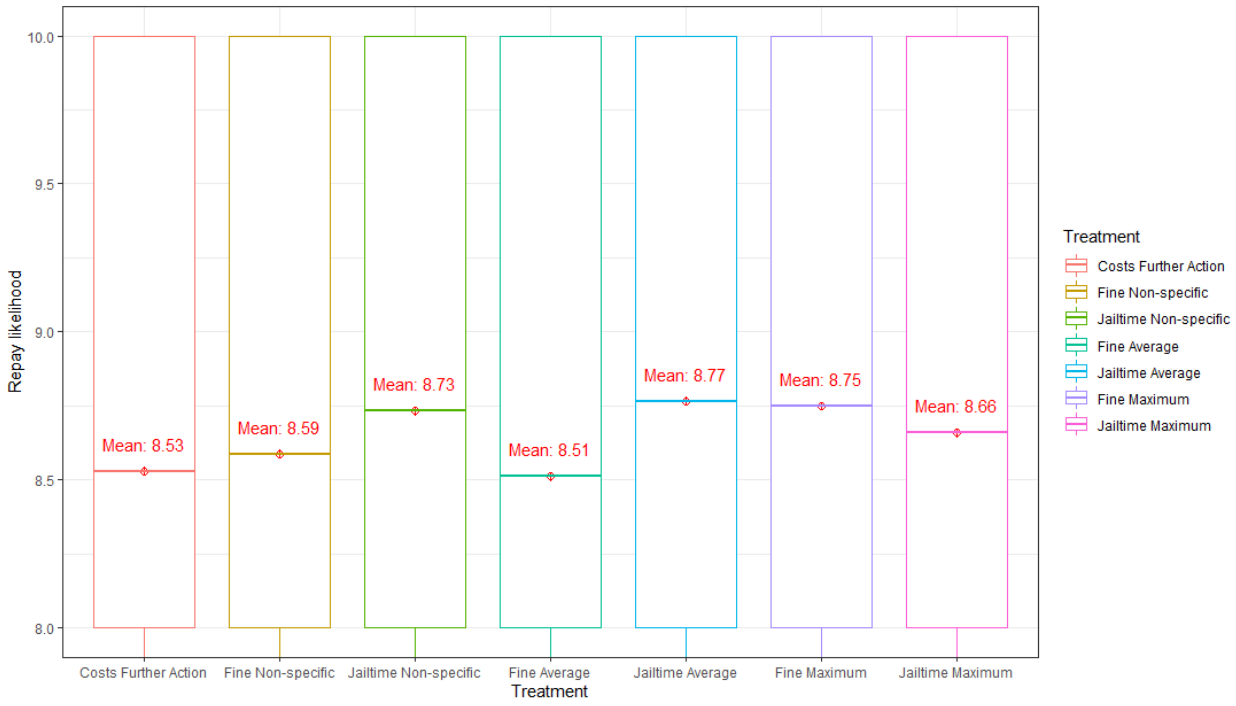


Table 14: OLS regression on paying tax - Experiment III

	(I) Pay tax – Full sample	(II) Pay tax – Full sample	(III) Pay tax – Full sample	(IV) Pay tax – Full sample
Fine Non-specific	0.058	0.055	-0.063	-0.059
	(0.082)	(0.080)	(0.075)	(0.073)
Jailtime Non-specific	0.203*	0.204*	0.208**	0.209**
	(0.082)	(0.080)	(0.075)	(0.074)
Fine Average	-0.019	-0.021	0.051	0.036
	(0.082)	(0.080)	(0.075)	(0.073)
Jailtime Average	0.237**	0.251**	0.364***	0.372***
	(0.082)	(0.080)	(0.075)	(0.074)
Fine Maximum	0.220**	0.223**	-0.050	-0.042
	(0.082)	(0.080)	(0.075)	(0.074)
Jailtime Maximum	0.130	0.132	0.310***	0.304***
	(0.082)	(0.080)	(0.075)	(0.074)
Punishment Severity			0.214***	0.202***
			(0.011)	(0.010)
Realism			0.198***	0.192***
			(0.009)	(0.009)
Age		0.007***		0.007***
		(0.002)		(0.002)
Male		-0.334***		-0.241***
		(0.045)		(0.041)
Non-binary		0.066		0.121
		(0.260)		(0.238)
Risk Propensity		-0.131***		-0.105***
		(0.009)		(0.009)
Constant	8.530***	8.991***	5.777***	6.197***
	(0.058)	(0.101)	(0.089)	(0.119)
R^2	0.003	0.047	0.177	0.204
Adjusted R^2	0.002	0.046	0.176	0.203
N	7,064	7,017	7,020	7,017

* p<0.05, ** p<0.01, *** p<0.001.
Notes: Our dependent variable is how likely the participants are to pay their taxes on a 0-10 scale. Standard errors are presented in parentheses.

Table 15: Mediation analysis on paying tax - Experiment III

Mediator	Independent Variable (IV): Treatment	Effect of IV on mediator	Unique effect of mediator	Indirect effect	BC 95% CI	
					Lower	Upper
Realism	Fine Non-specific, Costs Further Action	0.714***	0.408***	**0.291**	**0.210**	**0.375**
		(0.096)	(0.018)	(0.042)		
	Jailtime Non-specific, Costs Further Action	-0.256	0.303***	**-0.077**	**-0.110**	**-0.046**
		(0.052)	(0.017)	(0.016)		
	Fine Average, Costs Further Action	0.047	0.344***	0.016	-0.006	0.039
		(0.034)	(0.017)	(0.012)		
	Jailtime Average, Costs Further Action	-0.263***	0.268***	**-0.070**	**-0.088**	**-0.054**
		(0.027)	(0.016)	(0.009)		
	Fine Maximum, Costs Further Action	0.024	0.295***	0.007	-0.005	0.019
		(0.020)	(0.017)	(0.006)		
	Jailtime Maximum, Costs Further Action	-0.214***	0.252***	**-0.054**	**-0.066**	**-0.043**
		(0.019)	(0.016)	(0.006)		
Punishment Severity	Fine Non-specific, Costs Further Action	-0.107	0.416***	-0.045	-0.115	0.024
		(0.085)	(0.021)	(0.036)		
	Jailtime Non-specific, Costs Further Action	0.238***	0.370***	**0.088**	**0.055**	**0.123**
		(0.043)	(0.020)	(0.018)		
	Fine Average, Costs Further Action	-0.138***	0.338***	**-0.046**	**-0.068**	**-0.026**
		(0.031)	(0.019)	(0.011)		
	Jailtime Average, Costs Further Action	0.099***	0.362***	**0.036**	**0.019**	**0.054**
		(0.023)	(0.019)	(0.009)		
	Fine Maximum, Costs Further Action	0.241***	0.395***	**0.095**	**0.077**	**0.115**
		(0.016)	(0.020)	(0.010)		
	Jailtime Maximum, Costs Further Action	0.061***	0.303***	**0.019**	**0.009**	**0.029**
		(0.016)	(0.018)	(0.005)		

* p<0.05, ** p<0.01, *** p<0.001.

Notes: Our dependent variable is how likely the participants are to pay their taxes on a 0-10 scale. Independent variable (Message) is manipulated at two levels: a message of interest vs. control ("Costs Further Action"). Standard errors are presented in parentheses. Significant indirect effects (i.e. when confidence interval does not cross zero) and corresponding bias-corrected 95% confidence intervals are presented in bold.

Table 16: Background characteristics - Experiment IV

Group name	Test phrase	N	Likelihood of repayment	Punishment severity	Age	% Female	Risk propen-sity
Costs Further Action	We told you recently that you have an outstanding tax obligation with the government. If you pay this debt now, you can stop us taking any further action.	18	8.17 (1.34)	6.11 (1.64)	21.78 (2.32)	38.89	4.44 (2.20)
Fine Non-specific	We told you recently that you have an outstanding tax obligation with the government. If you do not pay your taxes, you may be subject to a fine.	18	8.44 (1.50)	7.50 (1.30)	22.44 (2.64)	72.22	4.44 (2.01)
Jailtime Non-specific	We told you recently that you have an outstanding tax obligation with the government. If you do not pay your taxes, you may be subject to prosecution that results in time in prison.	18	9.11 (1.05)	7.58 (1.77)	23.11 (3.26)	78.95	4.26 (2.23)
Fine Average	We told you recently that you have an outstanding tax obligation with the government. If you do not pay your taxes, we could fine you 5% of your current outstanding tax obligation.	17	8.88 (1.45)	7.41 (2.85)	23.71 (6.77)	70.59	5.47 (2.92)
Jailtime Average	We told you recently that you have an outstanding tax obligation with the government. If you do not pay your taxes, you could face 6 months in prison.	15	9.27 (1.16)	8.00 (1.60)	22.73 (2.58)	66.67	4.33 (1.99)
Fine Maximum	We told you recently that you have an outstanding tax obligation with the government. If you do not pay your taxes, you could face a fine of up to 100% of the outstanding tax obligation and the possibility that the government seizes your personal assets (e.g. property).	15	9.13 (1.46)	8.33 (1.54)	21.67 (2.61)	53.33	3.68 (2.87)
Jailtime Maximum	We told you recently that you have an outstanding tax obligation with the government. If you do not pay your taxes, you could face up to 7 years in prison.	15	8.36 (2.27)	7.93 (1.49)	22.64 (2.10)	64.29	4.50 (2.62)

Notes: Standard deviations are presented in parentheses.

Figure 4: Repayment likelihood by treatment group – Experiment IV

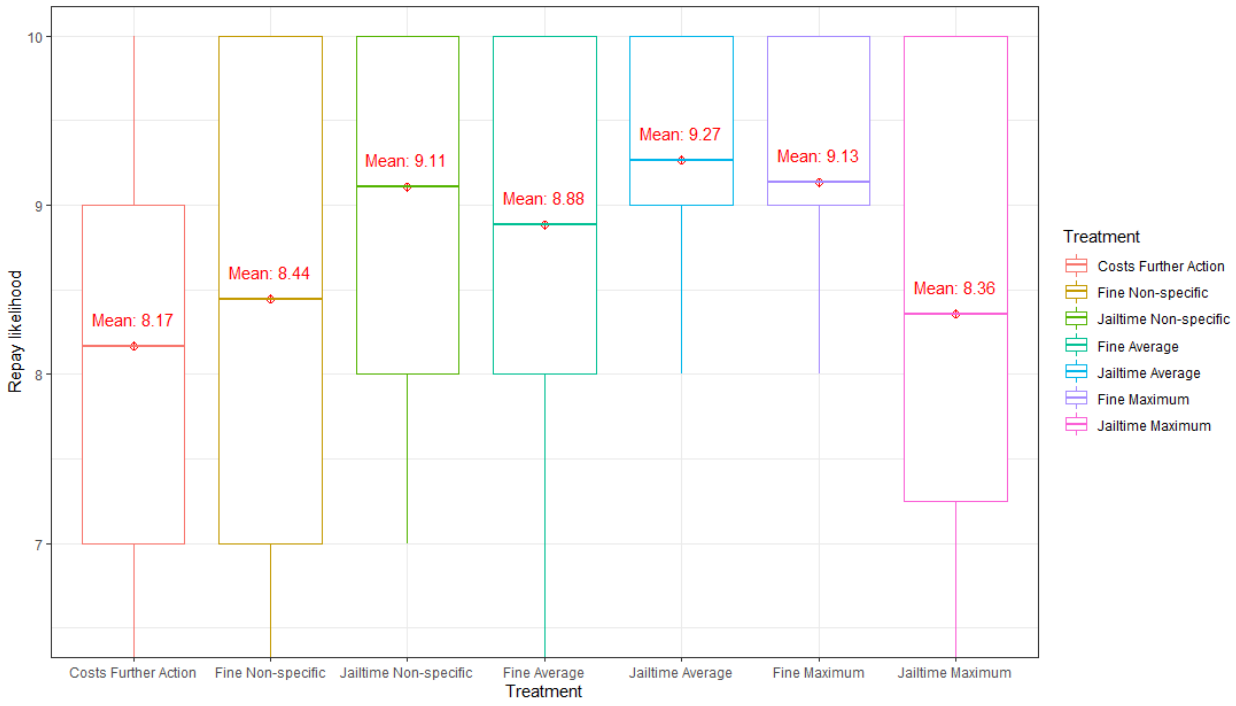


Table 17: OLS regression on paying tax - Experiment IV

	(I) Pay tax – Full sample	(II) Pay tax – Full sample	(III) Pay tax – Full sample	(IV) Pay tax – Full sample
Fine Non-specific	0.278	0.374		
	(0.493)	(0.506)		
Jailtime Non-specific	0.939	1.046*		
	(0.486)	(0.505)		
Fine Average	0.716	0.915		
	(0.500)	(0.534)		
Jailtime Average	1.100*	1.174*		
	(0.517)	(0.527)		
Fine Maximum	0.967	1.014		
	(0.517)	(0.523)		
Jailtime Maximum	0.190	0.258		
	(0.527)	(0.536)		
Fine group			0.633	0.746
			(0.408)	(0.421)
Jailtime group			0.771	0.855*
			(0.411)	(0.425)
Age		0.016		0.018
		(0.041)		(0.040)
Male		0.320		0.325
		(0.304)		(0.302)
Risk Propensity		-0.0003		-0.003
		(0.059)		(0.058)
Constant	8.167***	7.634***	8.167***	7.593***
	(0.349)	(1.031)	(0.350)	(1.018)
R^2	0.073	0.087	0.031	0.043
Adjusted R^2	0.022	0.009	0.014	-0.001
N	116	115	116	115

* p<0.05, ** p<0.01, *** p<0.001.

Notes: Our dependent variable is how likely the participants are to pay their taxes on a 0-10 scale. Standard errors are presented in parentheses. *Fine group* includes "Fine Non-specific", "Fine Average", and "Fine Maximum", while *Jailtime group* includes "Jailtime Non-specific", "Jailtime Average", and "Jailtime Maximum".

Table 18: Psychophysiological responses - Experiment IV

Group name	Pupillary Response, % change	EDA Change, % change	Fixation Count	Fixation Duration (sec)
Costs Further Action	17.72	12.17	251.28	55.06
	(16.39)	(24.75)	(68.81)	(16.09)
Fine Non-specific	21.66	19.54	266.39	62.00
	(25.75)	(19.99)	(44.22)	(11.71)
Jailtime Non-specific	30.12	56.21	234.32	56.64
	(21.55)	(38.89)	(75.60)	(18.89)
Fine Average	19.78	14.19	252.00	64.37
	(16.33)	(16.85)	(42.29)	(10.84)
Jailtime Average	44.20	57.31	267.80	59.67
	(20.24)	(55.29)	(44.41)	(7.62)
Fine Maximum	31.11	18.55	256.20	55.77
	(22.50)	(19.57)	(68.66)	(17.21)
Jailtime Maximum	18.38	3.74	273.86	56.01
	(17.35)	(32.62)	(31.98)	(7.95)

Notes: Standard deviations are presented in parentheses.

Figure 5: Pupillary response, % change by treatment group – Experiment IV

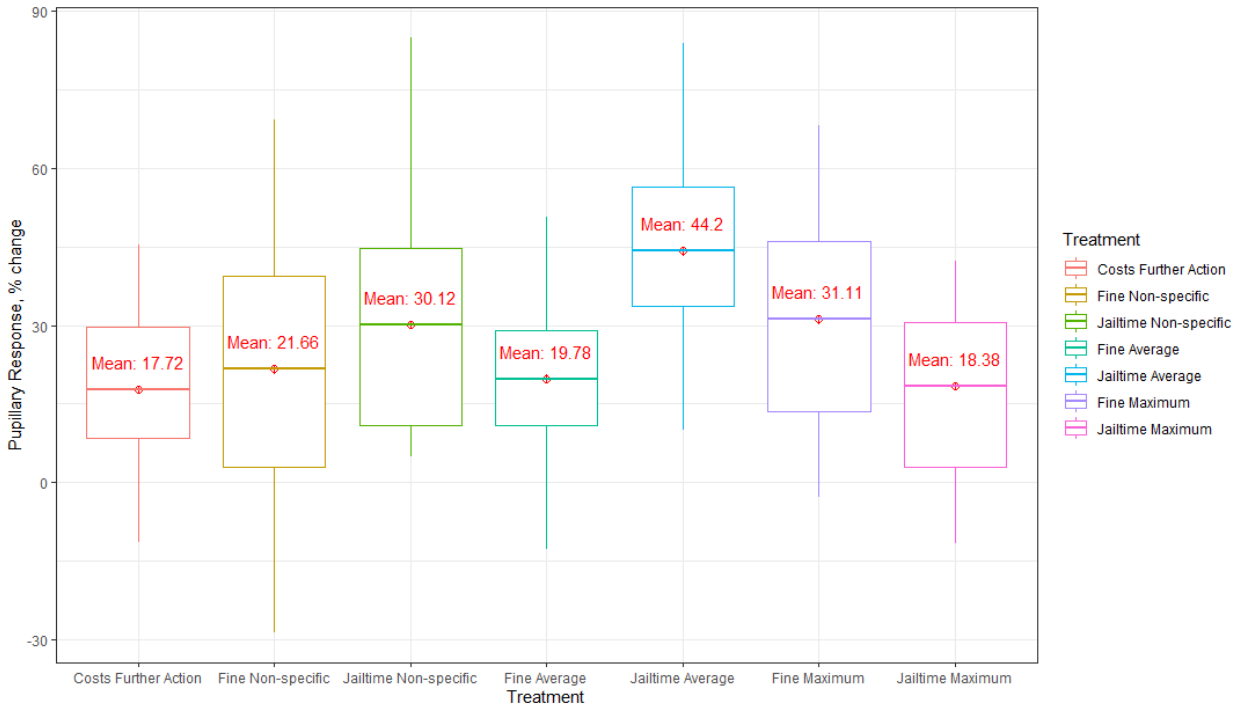


Figure 6: Skin Conductance Response (SCR), % change by treatment group – Experiment IV

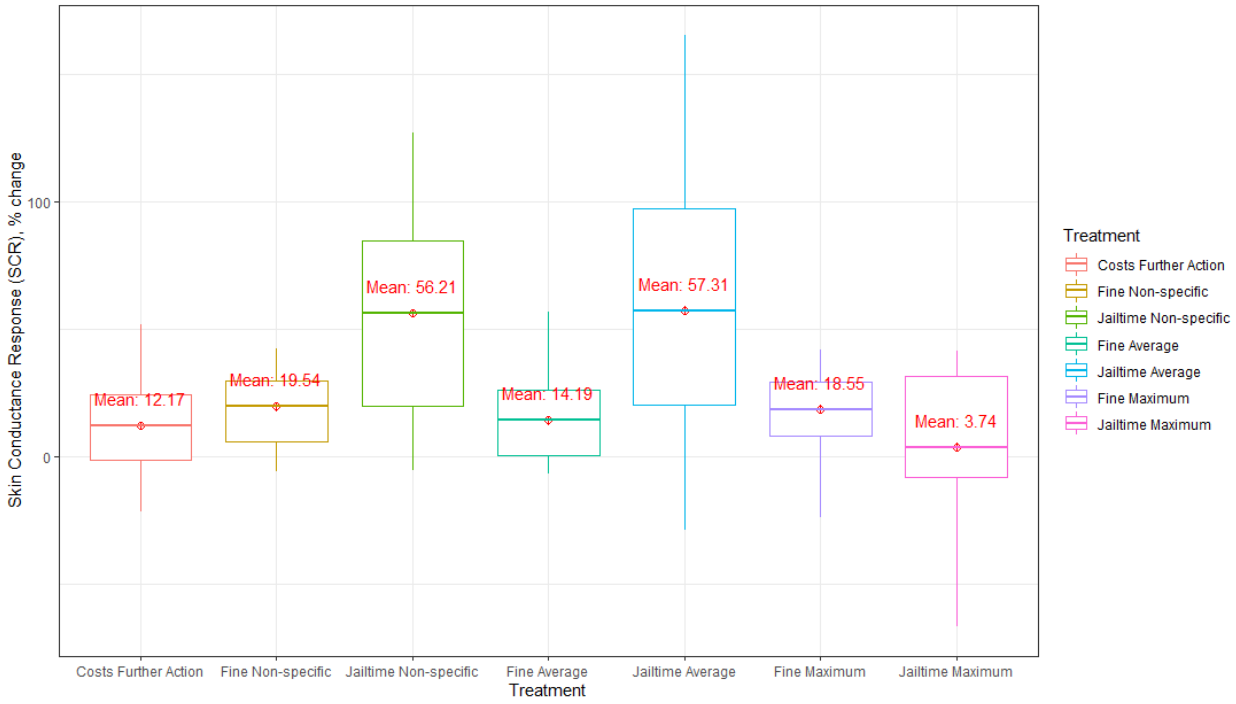


Table 19: Mediation analysis on paying tax - Experiment IV

Mediator	Independent Variable (IV): Treatment	Effect of IV on mediator	Unique effect of mediator	Indirect effect	BC 95% CI	
					Lower	Upper
Pupillary Response	Fine Non-specific, Costs Further Action	0.039	1.769	0.070	-0.228	0.379
		(0.072)	(1.106)	(0.146)		
	Jailtime Non-specific, Costs Further Action	0.062	-0.105	-0.007	-0.160	0.117
		(0.032)	(1.069)	(0.065)		
	Fine Average, Costs Further Action	0.007	1.956	0.013	-0.091	0.093
		(0.018)	(1.468)	(0.043)		
	Jailtime Average, Costs Further Action	0.066***	1.157	0.077	-0.109	0.252
		(0.016)	(1.247)	(0.090)		
	Fine Maximum, Costs Further Action	0.027	-1.480	-0.040	-0.165	0.029
		(0.014)	(1.284)	(0.094)		
	Jailtime Maximum, Costs Further Action	0.001	1.534	0.002	-0.040	0.056
		(0.010)	(1.972)	(0.022)		
	Fine group, Jailtime group, Costs Further Action	0.068*	1.532*	**0.104**	**0.004**	**0.244**
		(0.028)	(0.642)	(0.063)		
Skin Conductance Response (SCR)	Fine Non-specific, Costs Further Action	0.074	3.950***	0.291	-0.333	0.815
		(0.075)	(0.861)	(0.288)		
	Jailtime Non-specific, Costs Further Action	0.220***	2.054***	**0.452**	**0.209**	**0.817**
		(0.054)	(0.519)	(0.155)		
	Fine Average, Costs Further Action	0.007	2.823*	0.019	-0.134	0.142
		(0.024)	(1.046)	(0.067)		
	Jailtime Average, Costs Further Action	0.113**	1.645**	**0.186**	**0.067**	**0.371**
		(0.036)	(0.469)	(0.077)		
	Fine Maximum, Costs Further Action	0.013	1.773	0.023	-0.044	0.091
		(0.016)	(1.081)	(0.033)		
	Jailtime Maximum, Costs Further Action	-0.014	2.560*	-0.036	-0.162	0.047
		(0.017)	(1.079)	(0.052)		
	Fine group, Jailtime group, Costs Further Action	0.166***	1.487***	**0.247**	**0.099**	**0.424**
		(0.046)	(0.371)	(0.084)		

* p<0.05, ** p<0.01, *** p<0.001.

Notes: Our dependent variable is how likely the participants are to pay their taxes on a 0-10 scale. Independent variable (Message) is manipulated at two levels: a message of interest vs. control ("Costs Further Action"). Standard errors are presented in parentheses. Significant indirect effects (i.e. when confidence interval does not cross zero) and corresponding bias-corrected 95% confidence intervals are presented in bold.

Appendix: An example of the control letter in Experiment I

Dear Sir/Madam

Date of issue 20 February 2012
Reference REFERENCE NUMBER

Please pay £999999999999.99

Our records show that your Self Assessment tax payment is overdue.

Please call **0845 366 7809** now. You can pay by debit card, credit card, or Direct Debit. You can also pay using internet and telephone banking. For more information on how to pay, go to **www.hmrc.gov.uk/payinghmrc**

If you don't believe that this payment is overdue, please contact us on the number above.

If you have already paid, thank you. If not, please act now.

Yours faithfully

Officer of Revenue and Customs

IDMS99P HMRC 10/10

Appendix: An example of the public services letter in Experiment I

Dear Sir/Madam

Date of issue 20 February 2012
Reference REFERENCE NUMBER

Please pay £999999999999.99

Our records show that your Self Assessment tax payment is overdue.

According to a 2009 opinion poll, our most valued public services are the NHS, schools, care for the elderly, and the police. Every single tax payment helps to run these services.

Please call 0845 366 7809 now. You can pay by debit card, credit card, or Direct Debit. You can also pay using internet and telephone banking. For more information on how to pay, go to **www.hmrc.gov.uk/payinghmrc**

If you don't believe that this payment is overdue, please contact us on the number above.

If you have already paid, thank you. If not, please act now.

Yours faithfully

Officer of Revenue and Customs

IDMS99P HMRC 10/10

Appendix Table A1: Regression on predictors of treatment group membership, Experiment I

	(I) Male	(II) Age	(III) Initial debt	(IV) Self-employed	(V) Accountant
Reminder	-0.012	-0.417*	183.765*	-0.005	-0.004
Moral Duty	0.009	-0.098	126.878	-0.007	-0.001
Moral Equity	-0.033	-0.141	35.964	-0.009	-0.008
Public Services	-0.514	0.050	66.959	-0.003	-0.004
Costs & Gain	-0.008	-0.365*	169.405	-0.002	-0.007
Costs & Loss	-0.031	-0.556	56.983	-0.004	-0.000
Deficit	-0.016	-0.279	29.585	-0.002	0.002
Progress	-0.040	-0.979	126.878	-0.002	-0.001
Progress & Thanks	-0.021	0.126	17.851	-0.007	0.004
Monitoring	-0.026	-0.437*	-74.332	-0.013**	-0.008
Monitoring Day	0.045	0.204	31.245	-0.003	0.001
Monitoring & Norms	-0.054	-0.001	68.738	-0.004	0.003
Moral & Monitoring & Norms	-0.016	-0.144	123.427	-0.008	0.005
N	119,728	119,309	119,728	119,728	119,728

* p<0.05, ** p<0.01, *** p<0.001.

Notes: The Control group is the omitted variable. The standard errors are omitted owing to space constraints. Regressions (I), (IV) and (V) are logistic, and regressions (II) and (III) are OLS.

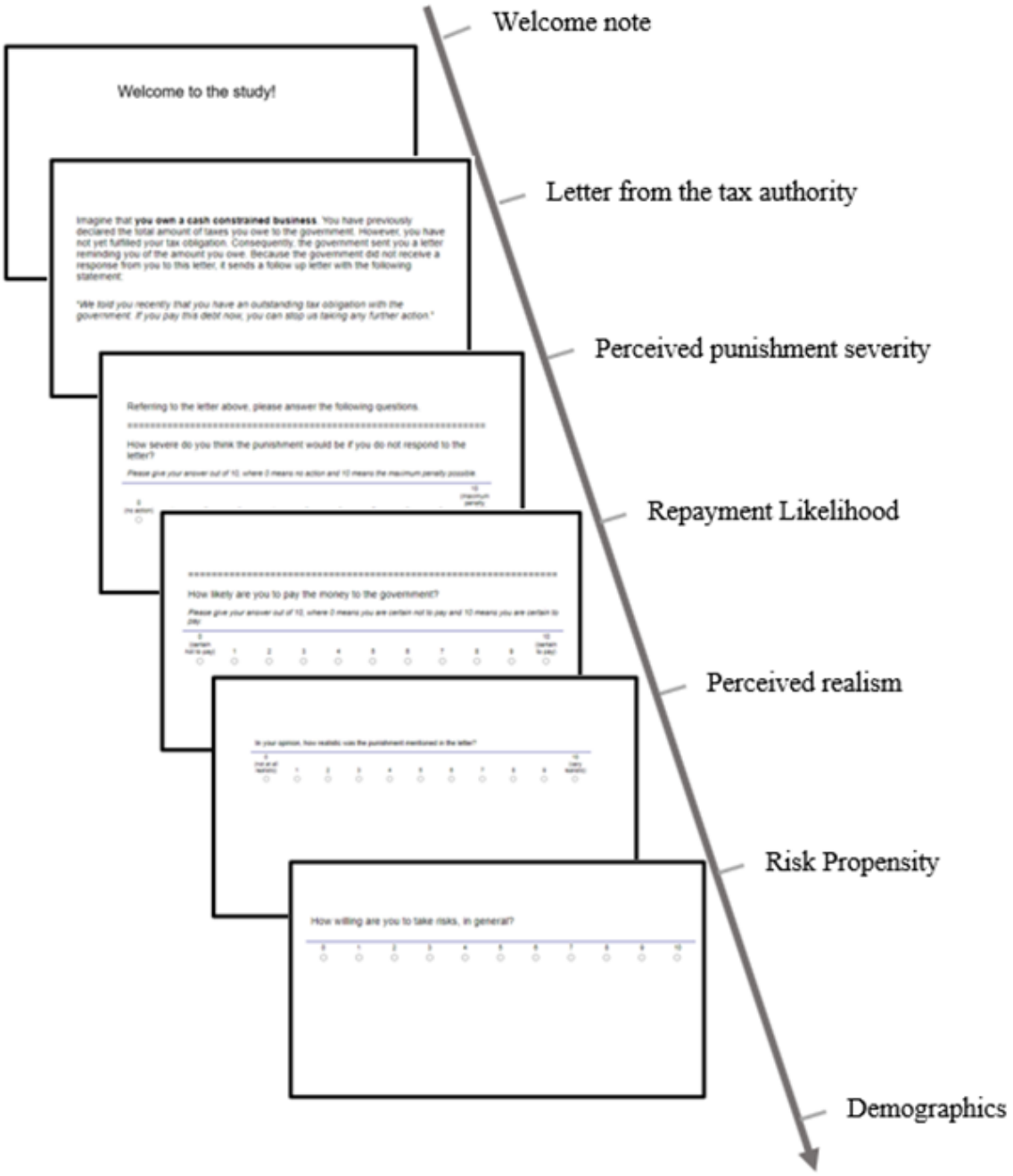
Appendix Table A2: Regression on predictors of treatment group membership, Experiment II

	(I) Male	(II) Age	(III) Initial debt	(IV) Self-employed	(V) Accountant
Interest Today	0.007	-0.002	4.321	-0.001	-0.002
Interest Tomorrow	-0.007	0.004	-29.939	-0.003	-0.004
Less Effort	0.001	-0.381**	54.453	0.001	0.003
More Effort	0.003	-0.212	30.573	-0.003	-0.009
Costs Further Action	-0.009	-0.314	27.723	0.001	0.003
Costs Peace	0.006	-0.226	-11.724	0.001	0.001
Interest & Effort	0.002	-0.390**	32.336	-0.059*	-0.010*
Interest & Action	0.006	-0.125	-15.301	0.002	0.000
Contact Now	0.002	-0.325**	46.167	-0.001	-0.003
Contact Now Information	0.001	-0.381**	-11.741	-0.003	-0.008
Contact Later Information	0.003	-0.286*	14.039	-0.002	-0.002
N	204,936	204,413	204,889	204,936	204,936

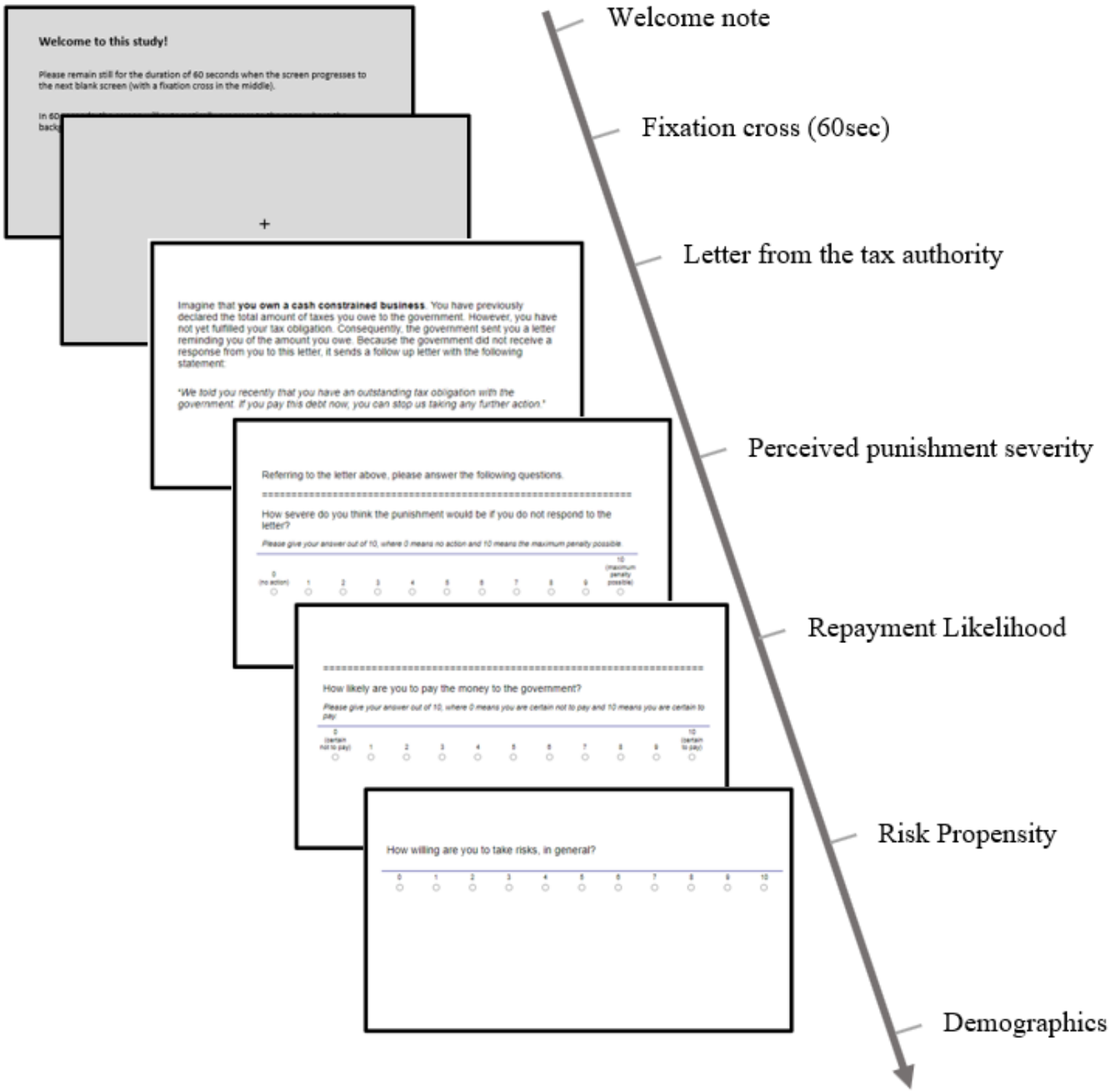
* p<0.05, ** p<0.01, *** p<0.001.

Notes: The Control group is the omitted variable. The standard errors are omitted owing to space constraints. Regressions (I), (IV) and (V) are logistic, and regressions (II) and (III) are OLS.

Appendix Figure A1: Experiment III Timeline



Appendix Figure A2: Experiment IV Timeline



Appendix Table A3: Technical specifications of the pupillometry and eye tracking measurements

Core parameters	Parameter description	Parameter specifications adopted in the experiment
Apparatus	Sampling procedure	Binocular recording procedure was used (i.e. pupil dilation and eye tracking measures are based on the data acquired from both left and right eyes of the participants)
	Name and produce of the eye tracking device	Tobii TX300, Tobii (Sweden)
	Type of eye tracking device	Desk-mounted
	Sampling rate	300 Hz
	Sampling rate variability	0.3%
	Processing latency	1.0 – 3.3 ms
	Accuracy[32]	0.4^0 – at ideal conditions[33], 0.3^0 - at 25^0 gaze, 0.6^0 - at 30^0 gaze, 0.6^0 – at 1 lux[34], 0.4^0 – at 300 lux, 0.5^0 – at 600 lux, 0.5^0 – at 1000 lux.
	Precision	0.01^0 – with Stamper filter
	Eye tracking software used	Tobii Pro Lab v 1.171.1
	Chin rest used	No
Monitor	Screen size	23‴
	Screen resolution	1920 x 1080 pixels
	Distance between participant and screen	Operating distance: 50-80cm Default distance used in this study: 65cm
Calibration	How many points in calibration	9-point calibration (regular calibration type selected)
	Amount of recalibration	Up to two recalibrations were performed during the initial calibration period preceding the main experiment[35]
Vision	Participant vision	Normal or corrected-to-normal.
Areas of Interest (AOIs)	AOIs used for pupillometry and eye tracking data analysis	Messages from tax authorities outlined in Table 16 and administered to participants via Tobii Pro Lab software were used as AOIs.
Event detection	Algorithm used for event detection	The IV-T fixation algorithm (Komogortsev et al., 2010) was adopted via the selection of global settings in the eye tracking software (Tobii Pro Lab v 1.171.1). A rather conservative 60ms threshold was selected within IV-T Tobii filter parameters to define eye fixations.

[32] The angular average distance from the actual gaze point to the one measured by the eye tracker.

[33] The default experimental setup of this study conforms to the definition of 'accuracy under ideal conditions' outlined in Tobii (2014) as follows: (i) the head movement of the participant is fixed in a chinrest; and (ii) data collected immediately after calibration, in a controlled laboratory environment with constant illumination, with 9 stimuli points (related to the 9-point calibration procedure undertaken in this study) at gaze angle $\leq 18^0$.

[34] Unit of illuminance and luminous emittance, measuring luminous flux per unit area. One *lux* is equal to one lumen per square meter.

[35] We used up to two recalibrations in cases when participants did not calibrate sufficiently well within at least one of the nine calibration points. Upon each calibration, we inspected the calibration and validation results as instructed in Tobii Pro Lab user manual (https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/Tobii-Pro-Lab-User-Manual/#page=49) and only proceeded to the main experiment when the calibration and validation results were satisfactory. Neither of the participants required more than two recalibrations.

Appendix: Theoretical Framework

In the standard model of tax evasion, the taxpayer faces a decision under risk, with the extent of evasion chosen to maximize expected utility (Becker, 1968; Allingham & Sandmo, 1972; Yitzhaki, 1974). The risk arises from the possibility that the tax authority will discover the tax evasion by conducting an audit. This model has often been used to assess how much income is *declared* to tax authorities (see Alm, 2012, for a review), but it is rarely used to understand the decision to *pay* the declared income.[36] Clearly, there are at least two stages to tax compliance. The first is to decide whether to evade. Once that decision is taken, in the second stage the individual decides to pay the declared tax on time, pay the declared tax late, or not pay the declared tax. Of course, in equilibrium the second stage reasoning affects the first stage decision, but we focus exclusively here on the second stage to provide a clear link to the natural field experiments.

In many countries, the costs of not paying declared income take the form of fines and/or interest on the outstanding tax liability. For instance, in the UK the penalty system is structured as follows. If payment is 30 days late, the agent must pay interest of five per cent on the tax that is owed at that date. If payment is six months late, the agent must pay five per cent of the tax that is owed at that date, in addition to the fine incurred at 30 days. If payment is twelve months late, the agent must pay five per cent of the tax that is owed at that date, in addition to the fines already incurred.[37] The structure is staggered because income tax is paid in instalments that are due every six months.[38]

We propose a model (with some inspiration drawn from Wang & White (2000) and Gross et al. (2013)) that attempts to describe a simple situation where: (a) individuals who owe taxes and are potentially liquidity-constrained (defined as earning less than a threshold level of income) have no margin on which the intervention can act; whereas (b) individuals who are not liquidity-constrained will trade off an explicit financial liability and a 'moral' cost of the type described in Levitt and List (2007), created through the use of descriptive norms, injunctive norms, public services appeals, and so on, with their outside option in the financial market and their inherent disutility from paying taxes right away versus waiting. We call this group the procrastinators, although we do not test procrastination *per se* in this paper (and other cost-benefit utility models could also be employed to test our hypotheses about the effect of moral costs).

[36] In contrast, countries such as the US and Canada explicitly refer to payment when defining compliance, see Boame (2008).
[37] There is an element of O'Donoghue and Rabin's (1999) model here where agents may have present-biased preferences in determining whether to act now or wait until later.
[38] http://www.hmrc.gov.uk/rates/interest-late-pay.htm; http://www.hmrc.gov.uk/sa/deadlines-penalties.htm#6.

As mentioned in the prior section, we assume nonstrategic play because all of the consumers have already disclosed their income. Agents experience a disutility from having to pay their tax today ($\beta > 1$), but do not experience the same disutility from the prospect of paying the tax in the future.[39] An agent faces a choice of whether to pay the tax now ($i = 0$) or pay the tax in the future ($i = 1$). Thus, the agent will seek to choose i to maximize the following utility function:

$$U = u_i(\mathrm{Y},\mathrm{M},t) = \begin{cases} \mathrm{Y} - \beta t & \text{if } i = 0 \text{ and } \mathrm{Y} \geq \beta t \\ \max[\mathrm{Y} + (1+r)t - \mathrm{M} - (1+\alpha)t, 0] & \text{if } i = 1 \text{ or } \mathrm{Y} < \beta t \end{cases}$$

Y = realized income
M = moral cost
t = tax liability

The model begins on the self-assessment day. Our representative agent earns a stochastic labor income of Y, drawn i.i.d. from probability distribution f(y). She subsequently informs the tax authority of her income which, through a deterministic rule, is converted into a tax liability which we call t, $t \geq 0$. We impose a no-borrowing condition so that our agent does not have the option of borrowing money at a prevailing market rate.[40] Thus, if our agent experiences a negative income shock ($\mathrm{Y} < \beta t$), luck has chosen her hand – she has no choice but $i = 1$. Otherwise, the agent has a choice of when to pay her tax. If the agent does not pay her tax immediately, interest fees accrue on her tax liability, growing geometrically at a rate of $\alpha \in (0,1)$. On the other hand, the agent can reinvest the funds at a rate of $r \in (0,1)$. Lastly, we assume that the agent faces a moral cost when not paying her tax liability immediately ($M > 0$), induced by a letter sent by the tax authority.

In order for her to prefer paying now (i.e., choose $i = 0$), it is sufficient that Condition 1 is met:[41]

$$\text{paying now} \succ \text{paying later} \iff \mathrm{Y} - \beta t > \mathrm{Y} + (1+r)t - \mathrm{M} - (1+\alpha)t$$

Condition 1 holds as long as $\beta + r - \alpha < \frac{\mathrm{M}}{\mathrm{t}}$. We introduce ϕ as the fraction of individuals that pay immediately. By defining individuals with a negative income shock in the first period ($\mathrm{Y} < \beta t$) as liquidity-constrained, we see that the above simple model generates the following prediction:

[39] This assumption is similar to the asymmetric impatience exhibited by agents with quasi-hyperbolic time preferences of the kind described in Phelps and Pollak (1968), Laibson (1997), and especially O'Donoghue and Rabin (1999).
[40] In this sense, the realized income is treated in a manner similar to "cash-on-hand" as in Deaton (1991). Explicit claims on future income cannot be used to cover the tax liability.
[41] While we use strict preference and inequalities throughout, the results would not significantly change if we instead used weak inequalities.

PREDICTION 1: Liquidity-constrained individuals, due to lack of access to lenders and low income, will pay later irrespective of whether they receive a reminder letter or not. More formally, $\frac{\partial \phi}{\partial \mathrm{M}} = 0$.

In addition, note that for individuals who are not liquidity constrained, Condition 1 will become easier to satisfy if the tax penalty rate increases or if the moral cost is increased, two policy levers that the tax authority can control. On the other hand, as the prevailing interest rate grows, as the size of the tax liability grows, and as the disutility of paying taxes today grows, Condition 1 becomes more difficult to satisfy, *ceteris paribus*.

PREDICTION 2*:* The fraction of individuals paying immediately is increasing with the tax penalty and moral cost and is decreasing with the interest rate (outside option), size of tax liability, and disutility of present tax payment, i.e. $\frac{\partial \phi}{\partial \alpha} > 0, \frac{\partial \phi}{\partial \mathrm{M}} > 0, \frac{\partial \phi}{\partial r} < 0, \frac{\partial \phi}{\partial t} < 0, \frac{\partial \phi}{\partial \beta} < 0$.

These two predictions focus our attention on the saliency of costs. We will talk about these channels in greater detail in the methods section of the field experiment.