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

Permanent WRAP URL:
http://wrap.warwick.ac.uk/88777

Copyright and reuse:

Publisher's statement:

A note on versions:
The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

The Behavioralist As Tax Collector:
Using Natural Field Experiments to Enhance Tax Compliance*

Michael Hallsworth
Imperial College London
The Behavioural Insights Team

John A. List
University of Chicago

Robert D. Metcalfe
University of Chicago

Ivo Vlaev
University of Warwick

November 30, 2016

Abstract

This paper presents results from two large-scale natural field experiments that tested the effect of social norm messages on tax compliance. Using administrative data from more than 200,000 individuals in the United Kingdom, we show that including social norm messages in standard reminder letters increases payment rates for overdue tax. This result offers a rare example of social norm messages affecting tax compliance behavior in a real world setting. We find no evidence that loss framing is more effective than gain framing. Descriptive norms appear to be more effective than injunctive norms. Messages referring to public services or financial information also significantly increased payment rates. The field experiments accelerated the collection of tax revenue at little cost.

Keywords: taxation, natural field experiment, social norms.
JEL: C93; D03; H26

1. Introduction

Understanding how to motivate individuals to pay their taxes has become a major issue for research in economics and public policy (Andreoni et al., 1998). The pioneers in this area were Allingham and Sandmo (AS) (1972), who adapted the Becker (1968) model to analyse the factors that influence the decision to evade tax. Their model predicts that tax evasion rates fall as the probability of detection and the degree of punishment increases. Various empirical approaches have been employed to test the AS model; notably, recent years have seen a surge in the use of field experiments (Blumenthal et al., 2001; Kleven et al., 2011; Ariel, 2012; Del Carpio, 2013; Carrillo et al., 2014; Gangl et al., 2014; Bott et al., 2014; Pomeranz, 2015; Castro & Scartascini, 2015; Dwenger et al., 2016).[1]

The great majority of studies have, however, focused on the initial decision to declare or disclose income to the government. This paper has a different emphasis. Rather than analysing the honest declaration of income, we use large-scale natural field experiments to examine the factors that influence timely *payment* of taxes. Such an exploration is relevant because an estimated 16% of the gross tax gap in the US is from enforced or late payment (Slemrod, 2007). Focusing on tax payments offers significant practical advantages over income declaration. For instance, rather than inferring compliance from changes in the level of income declared, there is a tighter causal link between intervention and behavior: a known tax amount is outstanding, a message requests payment, and the official tax record shows whether the payment has been made. Of course, our sample consists only of non-compliant taxpayers who initially declared their income to the government, so our results are only externally valid for this population of taxpayers.

The inspiration for our treatments comes from the growing literature suggesting that moral costs can be an important mechanism for changing human behavior (Levitt & List, 2007). While such a mechanism has been advocated in the theoretical literature (Postlewaite, 2010), there are few field experimental studies of its efficacy. We present two natural field experiments that test the impact of tax payment reminders featuring messages aimed at increasing moral costs.

The first natural field experiment we present was carried out with the United Kingdom tax administration authority (Her Majesty's Revenue and Customs (HMRC)) and the then UK Cabinet

---

[1] Hallsworth (2014) gives an overview of these field experiments and places the current study in the context of other similar projects. There are also field experiments that provide relevant findings, but not strictly in the field of tax compliance. Fellner et al. (2013) present a field experiment on increasing compliance with television licensing amongst Austrian citizens. Bhargava and Manoli (2012) conduct a field experiment with the U.S. Internal Revenue Service to increase take-up of benefits. There are many existing field experiments that deal with the effect of social norm information in wider policy areas, most notably energy consumption (Schultz et al., 2007). Our research also speaks to the literature that suggests that information provided by governments can affect citizen behavior (Chetty & Saez, 2009; Kling et al., 2011; Liebman & Luttmer, 2011).

Office's Behavioural Insights Team (BIT).[2] The experiment focused on those taxpayers who had already declared their income to be taxed, but who had not yet paid their tax liabilities.[3] We randomized five messages across around 100,000 individual taxpayers: three norm-based messages and two public services messages. A control group received a standard letter with no persuasive message of this kind. The results show that social norm and public services messages increased the likelihood of individuals paying their declared tax liabilities. The most successful moral cost message produced a treatment effect of 5.1% (0.1σ); this improvement was sustained until at least seventy days after the letters were issued. We find no evidence that a loss-framed message referring to public services is more effective than an equivalent gain-framed message. We estimate that £4.9 million was accelerated during the 23-day sample period due to the messages tested in the first field experiment.

These results led to a second large-scale natural field experiment (with the same government partners) that focused more carefully on message type. This second experiment investigated whether the most effective treatment from Experiment One could be replicated. In addition, we compared the effect of descriptive norms (i.e. what others do) and injunctive norms (i.e. what others think should be done). The experiment also interacted descriptive and injunctive norms. Finally, financial messages were included: they gave details of the added interest cost of non-payment and the payment vehicles that a person could use.

We find that descriptive norms have a significantly larger effect than injunctive norms on increasing payment rates. We also replicate the minority norm finding from the first experiment, which significantly increases the reliability of our results (Maniadis et al., 2014). We also find improvements from the financial messages (a 3.2-3.9 percentage point increase in payment rates). Many of the effects generated by the treatments remained economically meaningful and statistically significant seventy days after the letters were issued. This result suggests that social norm messages could have a persistent effect on behavior (see Allcott & Rodgers, 2014). Overall, we estimate that more than £9 million was accelerated during the initial 23 day sample period due to these messages. One should bear in mind that the marginal cost for the policymaker from this intervention was very small.

---

[2] It is important to note that the trials presented within this paper would not have happened without the concerted efforts of officials within the U.K. Government. In particular, the authors want to highlight the work of officials in HMRC and the BIT, with whom the authors worked to design and implement the trials, and collect the data upon which this analysis is based. Particular acknowledgement should go to Nick Down.

[3] The sample therefore consists of individuals who have exhibited non-compliant behavior. We speculate that these are individuals who may be less responsive to social norms around tax compliance (and thus the measured effects are lower bounds for a broader sample).

Overall, the results show that short messages referring to social norms and public services can influence the timing of tax payments. Many previous studies have concluded that the framing of messages does not matter – or, at least, that only sanction-based messages have an effect (Blumenthal et al., 2001; Kleven et al., 2011; Ariel, 2012; Hallsworth, 2014). There are many possible reasons for this difference (e.g., different countries, samples, timing, and incentives), and future research should explore what contexts allow moral costs to operate.

The remainder of the study is structured as follows. Section 2 gives a brief overview of the literature that informs the experiments. Sections 3 and 4 summarize the experimental design and results from the first natural field experiment. Sections 5 and 6 give the design and results for the second natural field experiment. Section 7 concludes.

2. Literature

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; AS, 1972; Yitzhaki, 1974). The risk arises from the possibility that the tax authority will discover and punish the individual's evasion. This model has often been used to assess how much income is *declared* to tax authorities (see Alm, 2012), but it is rarely used to understand the decision to *pay* the declared income. 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.[4] 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.

Various studies have proposed that giving information about the prevalence of compliance could increase the moral costs of non-compliance (Traxler, 2010; Myles & Naylor 1996; Frey & Torgler, 2007). The main mechanism proposed is that the actors internalize observed social norms, so that any deviation is accompanied by self-imposed costs such as feelings of guilt (Elster, 1989; Wenzel, 2004). There is some non-field experimental evidence that compliance increases after individuals are presented with such norm information (Wenzel, 2005a; Alm et al., 2013; Bobek, Hageman & Kelliher, 2013). Many field studies have, however, found no effect of providing social norm information; only recently has evidence emerged that social norms can affect compliance in real-world settings (Blumenthal, Christian & Slemrod, 2001; Slemrod, 2016).

---

[4] Countries such as the US and Canada explicitly refer to payment when defining compliance. See US Treasury (2009) and Boame (2008).

It is important to distinguish between different types of social norms. Cialdini et al.'s (1991) theory of normative conduct identifies 'descriptive' norms, which communicate the behavior of others, and 'injunctive' norms, which communicate the opinions of others. Put differently, descriptive norms say what others do; injunctive norms say what others believe, including what behaviors they approve of (see also Cialdini & Goldstein, 2004). Experimental research has shown that descriptive and injunctive norms have independent effects on behavior, strengthening the case for treating them as conceptually distinct (Rivis & Sheeran, 2003). The distinction between injunctive and descriptive norms has rarely been examined in the context of tax compliance; we believe that Bobek et al. (2007) and Bobek, Hageman & Kelliher (2013) are the only studies to have done so.

A related concept is that of moral duty. Rather than moral costs being increased because of information that the actor has deviated from the behavior of others, here the costs are generated by deviation from a moral standard (Erard & Feinstein, 1994; Reckers et al., 1994; Alm & Torgler 2011; Calvet Christian & Alm, 2014). From an empirical perspective, there have been survey-based (Feld & Larsen, 2012) and experimental (Bobek & Hatfield, 2003) studies that indicate beliefs about the morality of compliance can influence tax behavior. Again, however, other studies have shown no significant improvements (Torgler, 2004; Ariel, 2012).

Finally, we also included public services concerns, as introduced by Cowell and Gordon (1988), which allow taxpayers to derive utility from both income and public service provision. Most people realize that taxation is necessary to fund public services that benefit everyone, as well as specific items from which they gain in particular. Thus, there are at least two different approaches at work. First, a simple reciprocal mechanism of 'something for something', wherein tax is paid in order to receive some other kind of benefit (Fehr & Gachter, 1998). Second, a more altruistic perspective wherein paying tax is the means to create goods which everyone can potentially use and enjoy (Feld & Frey, 2010). There is some evidence that the level of benefits provided through taxation (and participants' awareness of these benefits) can influence compliance from both surveys (Spicer & Lundstedt, 1976) and experiments (Alm, McClelland & Schulze, 1992), but other studies have found no such effects (Torgler, 2004; Blumenthal et al., 2001).

Beyond simply introducing these sources of moral cost, we proposed that the way these concerns are framed can influence taxpayers' payment decisions. We focused particularly on the framing of social norms. For norms, one particularly promising approach is to make the norm more specific to the individual's situation or a group to which she belongs. Wenzel (2005b) has provided empirical support for this mechanism, reporting that perceived social norms affected tax

compliance only if the respondent identified with the group in question.

Such framings vary the level of abstraction or 'psychological distance' to the norm, as elaborated in *construal level theory* (Trope & Liberman, 2010), which has been shown to be important in laboratory experimental data. The theory postulates that we mentally represent, or 'construe', events or behaviors at different levels. Representations based on 'high-level' construals are experienced as psychologically distant and consist of a few abstract, decontextualized features that are more invariant than 'low-level' construals. In contrast, representations based on 'low-level' construals are experienced as more psychologically proximal – they are more concrete and consist of more contextualized here-and-now details. We hypothesized that reduced psychological distance (and therefore increased salience) would increase the moral costs incurred by the message.

We also tested the framing of public services concerns. While the discussion above focused on the benefits created through payment of tax, it is also possible to foreground the loss of benefits through non-payment. Although non-payment may lead to short-term gain for an individual, since such decisions are interdependent it could also contribute to a situation where all citizens eventually lose. Taxpayers could be loss-averse with respect to consumption and tax payments, so that events framed as losses are given disproportionate weight at the moment of choosing to pay or not (Tversky & Kahneman, 1992). There is evidence from other domains that this different 'goal/outcome framing' can influence behavioral outcomes (Meyerowitz & Chaiken, 1987; Ganzach & Karsahi, 1995; Rothman & Salovey, 1997). We therefore hypothesized that presenting the potential losses from non-payment of taxes will incur greater moral costs than presenting the potential gains from payment.

3. Experiment One: Methodology

3.1 Research setting

We collaborated in a natural field experiment on payment of taxes in the UK. Most income tax in the UK is collected by employers at source, through a "Pay As You Earn" system. Most individual taxpayers are therefore not required to submit a tax return. However, a variety of circumstances can require an individual to file a Self Assessment tax return, such as self-employment, multiple sources of income, or trustee status.[5] As Kleven et al. (2011) show, those who self-report income have the lowest levels of tax compliance.

Around ten million UK taxpayers participate in the Self Assessment scheme, which

---

[5] http://www.hmrc.gov.uk/sa/need-tax-return.htm - clearly this is not the same as withholding tax in the US.

requires them to file a return annually. Most participants also have to make two tax payments a year – the first by January 31, the second by July 31. If taxpayers do not pay the correct amount by these deadlines, then the tax authority has to pursue the debt. This process entails sending a reminder statement, followed by a combination of targeted letters and telephone calls. Ultimately, the tax authority has the power to enforce payment by seizing and auctioning goods and assets (Her Majesty's Revenue and Customs, 2010).

We incorporated the test messages into the letters sent to Self Assessment taxpayers who had not made the correct payment by July 31, 2011. All taxpayers had a debt of between £400 and £100,000 on August 1, 2011.[6] We excluded taxpayers with additional outstanding Self Assessment debts, since we could not rule out that they would not be receiving additional communications relating to their old debts. Scottish debts are subject to separate legal proceedings, and thus were not included. No other exclusion criteria were applied. These procedures resulted in a sample of 101,471 individuals geographically distributed across England, Wales, and Northern Ireland.

Letters were addressed solely to the individual who incurred the debt, and all were sent in standard HMRC envelopes. It is important to note that, due to administrative policy, letters had to be delivered to all agents who had not paid their taxes (including those in the control group). Recipients had no way of identifying the other participants, thanks to strict taxpayer confidentiality laws, so spillover effects are not a concern.[7] The letters themselves did not make reference to any experimental variation in wording. The individuals had not received any communications relating to these debts prior to the experimental letters. There were no promotional campaigns relating to the payment of tax debts during or prior to the issue of the letters. Thus, there is little identifiable risk of the results being contaminated by exogenous factors.

3.2 Experimental Messages

Table 1 summarizes our experimental design. We begin with a control letter, which has the basic HMRC reminder text. We contrast this baseline with five treatments. The experimental variation in these five treatments was simply the inclusion of a short phrase, in the standard typeface, after the first sentence. These messages were created to test the theories outlined in section 2. These phrases were constructed to persuade the recipient to pay the amount they owe,

---

[6] Debts below £400 and above £100,000 were subject to different actions and hence could not be included in the sample.

[7] Technically, it is possible that two taxpayers in a relationship may both incur self-assessment debts and may compare reminder letters. We consider the probability of this happening to be very small, and there is no evidence that it has happened.

and they fall into two main categories: those that refer to social norms, and those relating to the provision of public services. This experiment focused solely on the descriptive aspect of norms, and therefore the first message was a simple factual statement: "*Nine out of ten people pay their tax on time.*" Following the discussion of construal level theory above, this norm message was constructed to be as non-specific as possible. This 'basic norm' was therefore intended to capture a simple information effect aimed at correcting misperceptions.

The second phrase was identical to the basic norm, but explicitly associated the norm behavior with the United Kingdom: "*Nine out of ten people in the UK pay their tax on time.*" This variation was intended to increase the specificity (and therefore salience) of the norm message. Nationality was chosen as source of salience because Wenzel's (2004, 2005a) studies of Australian taxpayers found that levels of identification with Australia interacted with perceived norms of tax behavior to increase reported tax compliance. Torgler's (2007) work on 'tax morale' has also identified pride in one's country as an important factor in tax compliance. We hypothesized that sufficient numbers of taxpayers would identify with the UK, and hence find this formulation more salient, to produce a significant change in behavior compared to the generic norm.

The third phrase tested the effect of adding an explicit statement that the recipient is engaged in a minority behavior. Of course, this is the implicit conclusion of the basic and country norm messages, since all the reminder letters state that the recipient has not paid her tax yet. However, as Vlaev et al. (2011) show, decisions are made on the basis of even limited and superfluous information. The intent, therefore, was to assess the specific framing effect of presenting the recipient's behavior as a minority one. Again, we hypothesized that explicitly connecting the social norm to the recipient's own behavior would increase the salience of the norm message.[8] The 'minority norm' message read: "*Nine out of ten people in the UK pay their tax on time. You are currently in the very small minority of people who have not paid us yet.*"

The second strategy attempted to persuade individuals to pay taxes by focusing on the outcomes of taxation (i.e. public goods or services). We constructed a message that framed the issue of paying tax collectively, while also mentioning specific services that recipients were likely to have used themselves: "*Paying tax means we all gain from vital public services like the NHS [National Health Service], roads, and schools.*" As noted above, we are also interested in the

---

[8] It is worth noting that this additional statement can be seen as introducing two elements: a) an increased focus on the individual as such; b) the placing of that individual in a minority. In practice, it is difficult to communicate one concept without the other, which means we have not attempted to do so in this setting (the task of disentangling the two elements presents an opportunity for future laboratory-based research). It is debatable how substantial the effects of the first element are likely to be, since the opening sentence of every letter contains the phrase "*Our records show your Self Assessment payment is overdue*".

effects of framing the concept as a potential loss, so the following wording was introduced: *"Not paying tax means we all lose out on vital public services like the NHS, roads, and schools."* As in De Martino et al. (2006), the frame changes are limited to the minimum possible change in wording, thus limiting confounding factors.[9]

Two points relating to the messages should be noted. First, the control group receives the standard letter with basic information on size of debt and how to pay. Therefore, the main experimental treatments simply measure the change in behavior associated with the presence or absence of the test phrases. Second, it should be emphasized that any results generated in this trial show the effects of specific changes to message wording alone. Unlike many previous field experiments, the treatments were limited to relatively small changes to short phrases, rather than extensive changes to wording or a combination of phrases and visual content (cf. Blumenthal et al., 2001; see Hasseldine, 2000). At most, the phrases increase the letter length from 104 words to 133 words. Any interpretation of the effects should note the relatively modest nature of the treatments.

3.3 Sampling and randomization

The sample of 101,471 individuals was divided into six treatment groups to ensure that the study had adequate statistical power, capable of detecting effects roughly equal to a one percentage point difference in payment rates. Cases were subjected to a simple randomization process, with no blocking and equal allocation to each group (due to technical constraints). Each Self Assessment taxpayer has a unique reference number, which is created by computer-generated randomization syntax.[10] In order to select cases, the six different messages were assigned to 54 ranges of these reference numbers. Ranges were used because the technical ability did not exist to allocate taxpayer numbers to messages on a case-by-case basis. Since the taxpayer numbers were randomly generated, their contiguity was not considered a risk to randomization. Due to the fact that the letter ranges were inputted each day manually, 54 ranges represented the point at which the risk of implementation errors through complexity began to outweigh the marginal benefits.

---

[9] It should be noted that the public service messages refer to "paying tax", rather than "paying tax on time", as in the norm messages. To standardise one or the other would cause practical problems. If the social norm messages were to refer to "paying tax" instead, this would complicate the message considerably, since the proportion who pay *at all* is much higher than nine out of ten. Either the "nine out of ten" message would need to be changed to a percentage, which we considered to be less compelling (an assumption that was tested empirically in Experiment Two), or it would risk understating compliance. If the social norm messages were to refer to "paying tax on time", then the message would have to be diluted to something like "not paying tax on time weakens vital public services", or similar. Our judgment was that this was a less compelling and – more importantly – less defensible statement.

[10] Analysis conducted by HMRC revealed no significant correlations between taxpayer reference numbers and observable characteristics such as age and gender, which gave us further confidence in this process.

The resulting groups were similar in size, total value and mean value of debts, as well as mean taxpayer age. Aggregated figures also showed similar gender allocations across groups (see Table 1).

After the trial was conducted, we were able to obtain line-by-line debt data and run regressions to establish whether membership of a treatment group was significantly predicted by any of these variables (age, gender, and size of debt). We do observe some imbalance on the debt and gender variables, with a high degree of certainty in some instances (see Table A1). However, in the Results section we note that the coefficients do not change significantly with the introduction of covariates, and that there is a consistent pattern of results between the two experiments. The robustness of the results to controls gives us confidence in their reliability; the imbalances are an example of the inherent challenges of working with existing government processes in a live policy environment. We obtain balance across time with respect to the timing of letters received ('early' versus 'late', as explained below).

Another consideration in the sampling was that the volume of cases required the letters to be issued over five sequential days. To prevent the day of issue from creating any systemic variation, a Latin Squares design was used to ensure that an equal number of reference number ranges were allocated to each day (see Table 2).

3.4 Empirical Strategy

Our empirical strategy is to examine the effect on tax paying behavior of: (a) the impact of reminder letters as such; and (b) messages contained in those letters. Our dependent variable is whether the letter recipient made a payment to the tax authority, according to official tax records. As noted previously, letters had to be delivered to all agents who had not paid their taxes, including those in the control group. Therefore, to measure the effect of the baseline (reminder) letter, we staggered the issue of the letters over five days in August 2011. Since the date of issue was randomized, we can compare the compliance rates of those receiving the earliest letters versus the latest letters. If we do this at a point when the early letters have been received, but the late letters have not, we can estimate the effect of a receiving a reminder letter *per se*. Therefore, we compare the payment rates at August 23, 2011 of those who were issued letters on August 16, 2011 with those who were issued letters on August 22, 2011 (see timeline below).

Date	Action
16th August 2011	(n) EARLY letters sent

17th August 2011	
18th August 2011	
19th August 2011	EARLY letters received
[weekend]	
22nd August 2011	(n) LATE letters sent
23rd August 2011	Comparison point
24th August 2011	
25th August 2011	LATE letters received

We make this comparison using the following logit regression:

$$P_{i8} = \alpha_i + \beta_1 L_i^{early} + \varepsilon_i \tag{4}$$

where P_{i8} is whether the individual pays their debt in the first eight days (i.e. by August 23, 2011, which is before individuals in the 'late' group receive their letter), and L_i^{early} is a dummy variable that is 1 if the individual is randomized to receive the tax letter in the early period as opposed to the late period. We present the marginal coefficients in the summary results table, where β is the impact of being in the group receiving an early letter versus a late letter on tax payment.

To examine the impact of the messages we run the following logit regression:

$$P_{i23} = \alpha_i + \beta_1 L_i^{basic} + \beta_2 L_i^{country} + \beta_3 L_i^{minority} + \beta_4 L_i^{gain} + \beta_5 L_i^{loss} + \varepsilon_i \tag{5}$$

where P_{i23} is whether the individual pays their tax debt in the first 23 days.[11] L_i^{basic} is a dummy variable for the basic norm treatment, $L_i^{country}$ is a dummy variable for the country norm treatment, $L_i^{minority}$ is a dummy variable for the minority norm treatment, L_i^{gain} is a dummy variable for the gain-framed public services treatment, and L_i^{loss} is a dummy variable for the loss-framed public services treatment. In the results tables we present marginal effects of the logit regressions.[12] We also conduct a survival analysis in order to determine whether we obtain similar results using a different approach. The survival analysis methodology is summarised in the Appendix.

4 Experiment One: Results

Table 3 summarizes the impact of early versus late letters, in order to estimate the effect of receiving a reminder letter as such. The first two rows in Table 3 show that the impact of the

---

[11] We measure the letter effects during the first 23 days for the following reason. Given that the test messages were only included in the first letter sent by the tax authority, the most accurate point at which to measure their effects is immediately before the subsequent reminder letter is received. After factoring in the potential variation in postal delivery times, we calculate that the earliest these subsequent letters can arrive is 23 days after issue of the first letter.

[12] We also ran OLS estimates with robust standard errors and find identical coefficients and standard errors.

control reminder is around 6.7 percentage points - that is, those receiving a letter were nearly four times more likely to pay tax at this point than those who did not receive a letter.[13] Subsequent rows show the differences for early versus late letters for our five letters with additional messages. We find that the reminder letters with the norm framings have a 7.1-7.8 percentage points effect on payment. The public services letters also had an impact on payment, with an effect of between 5.9 and 7.6 percentage points. When we average across all the groups, we find that the effect of the reminder letters is 7.1 percentage points. This is not a precise estimate of the reminder effect, and is probably a lower bound estimate, because we could not observe whether additional payments were made after everyone had received a tax letter.

Result 1: Reminder letters accelerate tax payments.

We turn next to the impact of the messages. Figure 1 plots the percentage of people per day who pay in the first 23 days in each of the six treatment groups. The days for which there is no recorded payment are weekends.[14] From visual inspection, one can clearly see that differences emerge from August 25th 2011. In fact, most of the treatment groups have higher per day payment rates from this date than the control group.

The regression outputs in Table 4 show that the trial letters significantly increased the likelihood of payment occurring. These results were maintained after data on the taxpayer's age, gender, and size of debt were added to the model as linear covariates. The average effect for the basic norm is 1.3% (0.025σ) and for the country norm it is 2.1% (0.040σ). The effect size for the minority norm is much larger, with a payment rate 5.1 percentage points higher than the control (0.11σ). This minority norm effect is around 70% of the reminder letter effect. This minority norm effect is significantly greater than both the basic norm and the country norm (both $p<0.001$). The country norm is not significantly greater than the basic norm (diff=0.017, $p=0.26$), which does not support the construal level theory outlined in section 2. Both the gain- and loss-framed public services messages had an effect of 1.6% (0.030σ), which does not support the loss-framing hypothesis in section 2.

Results from the supplementary survival analysis confirmed that there was a significant difference in payment rates between the treatment groups and the control for a) the period up to 23 days; b) the period between 23 days and 48 days; c) the study period as a whole. We do not

---

[13] This is a similar effect size to Fellner et al. (2013) in the case of television license payments.
[14] August 29th 2011 was also a public holiday in the UK, which has clearly affected the payment data.

observe a statistically difference between treatment groups and control if the period after 48 days is analysed on its own (see Appendix).

Result 2: Letters with social norm messages accelerate tax payments, especially when the minority norm frame is used.

Result 3: Letters with public services messages accelerate tax payments.

Result 4: Loss-framed public service messages are not significantly more effective at increasing tax payments than gain-framed public service messages.

Thus, we find that all of the test messages have a positive impact on paying tax. Jointly, these five test messages increase payments after 23 days by 2.2 percentage points compared to the control message (p<0.000). In specification (III), we have included day fixed effects for when the letter was sent out. If we find that people are more likely to respond to later letters, then there might be a liquidity constraint argument to be made. Including the day effects makes no difference to the treatment coefficients. It is interesting to observe that the third and fifth days the letters were sent out had an increased payment effect of 1.2% and 3.5% (both p<0.01).

Although the 23 day period is most appropriate to isolate the experimental effects, we can extend this period to understand the persistence of these effects, a question that has recently attracted attention in other policy areas (see Rogers & Frey, 2014). Columns (IV) and (V) of Table 4 present payment rates at 48 and 70 days respectively. These periods were selected because they represented the points at which subsequent stages of the tax authority's debt recovery process were concluded. To clarify: 23 days represents the last point before the second letter was received by taxpayers. A third letter was then issued, arriving at 35-38 days. 48 days represents the last point before the debts were sent to a call center, where multiple calls were made to try to collect the debt. 70 days represents the last point before the debts were sent for a personal visit from an officer with the power to seize personal goods for payment. These results show that the country and minority norm messages remain economically meaningful and statistically significant at 48 days, and the minority norm message at 70 days (see Figure 2). The results therefore mean that more costly and intensive debt collection procedures were avoided.[15]

---

[15] HMRC were unable to provide us with reliable estimates of the cost of these actions, and we do not believe these figures have been made public through other channels.

We interacted the letter treatments with age and gender. None of the interactions were significant at the five percent significance level, except that men responded 2.4% more to the loss-framed message than women, which reflects earlier experimental evidence on framing effects, gender and taxation (Hasseldine & Hite, 2003). In Table 5 we present the results by quartiles of debt size, while also isolating those with debts in the top 5% and 1% of the distribution. It appears that the effect sizes are relatively similar for these various groups, but it is worth noting that the effect of the public services messages appears to be significantly larger for those with larger debts: for those in the top 1% of the distribution, the loss-framed public service message increased the payment rate by 17.6 percentage points.

We also examined whether these results occurred because the messages increased the recipients' perceptions of the costs associated with non-compliance. To do this, we ran a subsequent laboratory experiment where participants were randomized to view seven of the treatment letters (see Appendix). We measured participants' responses on five dimensions. The first two measures captured the two main elements of the AS model: likelihood of detection and severity of punishment; these were complemented by a more general measure of "threat" posed by the messages. The final two measures concerned how specific a participant thought the message was to them (to address the construal-level hypothesis) and the participant's self-reported likelihood of paying.

We take the country norm message as our comparison group because, of the seven messages, it produced the smallest effect on behavior. We found that there were no significant differences between the country norm and other treatment groups in likelihood of payment, likelihood of detection, specificity to recipient, or likelihood of payment. The general threat perception was significantly higher than the country norm for four of the six other messages: minority norm, minority status, local descriptive norm, and local and debt descriptive norm. Arguably, this presents weak evidence that increased perception of threat may be a factor contributing to these results, although the lack of significant differences in similar measures should also be noted.

We use the estimates from Table 4, Regression (I) to estimate the revenue accelerated from the messages in the 23 day period. The minority norm message generated £1.8 million in collections that would not have been present at the 23rd day otherwise.[16] If this approach had been taken on the whole sample, it would have accelerated £10.9 million at this point. If we take the

---

[16] This is calculated by examining the average effect of 3.8% on the minority norm group liabilities. So with 16,515 individuals in this group with a mean debt value of £2,810.51, a 3.8% increase in payment would be equal to (16,515* 0.038)*£2,811.

other estimates at face value, we calculate that £4.9 million otherwise would not have been paid in the experimental period.[17] We investigated the possibility of assessing the benefits from the trial by calculating the cost savings to HMRC from having to carry out enforcement activities on fewer debts, but it was unable to supply us reliable cost estimates.

Result 5: £4.9 million was accelerated in the first 23 days from the test messages.

5 Experiment Two: Methodology

5.1 Research setting

The second natural field experiment had the same setting as Experiment One, but was conducted a year later. The three main objectives of the second experiment were to examine the reliability of the minority norm effect, compare the relative effects of descriptive and injunctive norms, and assess the impact of including financial information. Letters were sent to Self Assessment taxpayers who had not made the correct payment by July 31, 2012. All taxpayers had a debt of between £351 and £50,000 on August 1, 2012.[18] As before, taxpayers with additional outstanding Self Assessment debts were excluded, which resulted in the sample containing 119,527 individuals from England, Wales, and Northern Ireland.

5.2 Sampling and randomization

The sample of 119,527 individuals was split into fourteen groups, resulting in a mean sample size of 8,538. While increasing the number of groups limited our ability to detect differences of less than two percentage points, it allowed more sophisticated comparisons between norm effects. As before, technical limitations meant we were unable to block cases in advance of the letters being issued. We retained the system of using taxpayer unique reference numbers as the units of randomization, but developed the procedure further. We identified 84 ranges of taxpayer reference numbers, reflecting the fourteen groups of letters that were to be issued over six days. We then used computer-based randomization to allocate these ranges to one of the treatment groups and to a particular day of issue. The resulting treatment groups were similar in size, total value, mean value of debts, mean taxpayer age, and gender distributions (Table 6).

---

[17] We discuss the welfare implications of these results in the Discussion.

[18] The value ranges therefore differ from the first experiment. This is owing to developments in the tax authority's procedures. Given the distribution of values, the impact of these changes on the size of our sample is relatively small.

Again, we ran a logistic regression to investigate whether membership of a particular group is significantly determined by these variables. When accounting for age, gender, size of debt, and use of an accountant, across our fourteen groups we found that in three of the 52 comparisons balance was a significant factor at the 0.05 significance level. The fraction injunctive norm and general descriptive norm groups were less likely to be male, and the minority status group were less likely to use an accountant.

As with Experiment One, letters were issued over sequential business days, although in this experiment six days were required to issue the letters. Rather than a Latin Squares design, we used the additional randomization step described above to allocate cases to be issued letters across the six days of the experiment (see Table 7).

5.3 Messages

As with Experiment One, letters were sent to all taxpayers in the sample. Letters similarly contained basic information about the size of debt and means of payment, but this information was shorter than in the previous experiment. Again, the experimental variation was the inclusion of a short phrase after the first sentence, this time in bold typeface. All other aspects of the treatments remain identical.

The thirteen messages in the second natural field experiment can be divided into three groups. First, six social norm messages that represented varying levels of psychological distance, including a replication of the minority norm message from Experiment One. Second, five messages constructed to measure the impact of injunctive, rather than descriptive, social norms (Cialdini et al., 1991). Third, two 'financial information' messages specifically related to the payment decision, namely the inclusion of additional payment information and a warning of interest charges.

The first group of messages varied psychological distance by making the norm more specific in two respects: in terms of geography, and in terms of the tax debt. An adapted form of the country norm message from Experiment One was used as a general descriptive norm message (*"The great majority of people in the UK pay their tax on time"*).[19] In a second message, the mention of the country was replaced with a reference to the recipient's local area, thereby increasing geographic specificity (*"The great majority of people in your local area pay their tax*

_____

[19] We adapted the country norm, rather than the basic norm, on the basis that it was more likely to produce a significant difference from the control condition. This was a relevant point, given the reduced power of Experiment Two compared to Experiment One. The general descriptive norm is very similar to the country norm message in Experiment One, except that it states 'the great majority', rather than 'nine out of ten'. This is because it could not be guaranteed that the 90% payment rate occurred in every *local* area in England, Wales and Northern Ireland (as stated in the local descriptive norm message), and therefore such a statement could prove incorrect if questioned.

on time"). Debt specificity was increased by a third message, which indicated the social norm for those with similar debts ("*Most people with a debt like yours have paid it by now*").

A fourth message combined the preceding two messages ("*The great majority of people in your local area pay their tax on time. Most people with a debt like yours have paid it by now*.") As with the movement from basic to country norm in Experiment One, these messages were intended to examine the effect of making a norm increasingly specific to a recipient. Geography and type of debt were identified as two dimensions that produced messages that remained applicable, accurate, and acceptable to recipients, even as specificity increased.

As noted above, we also included the message that produced the largest effect in Experiment One, namely the minority norm ("*Nine out of ten people in the UK pay their tax on time. You are currently in the very small minority of people who have not paid us yet*"). We hypothesized that this effect would be replicated in Experiment Two. However, Experiment One did not include the minority phrase – the second half of the message – as a separate message. We therefore did so in Experiment Two, in order to assess the specific effect of referring to minority status ("*You are currently in the very small minority of people who have not paid us yet*").

The second group of messages were constructed to assess the effect of messages based on injunctive norms. Experiment One showed that descriptive norm statements increased tax payments, raising the question of whether injunctive norm statements about attitudes towards taxpaying could have similar effects. Given existing evidence, we hypothesized that the injunctive and descriptive norms would have a significantly different effect on behavior – although we did not propose a direction for this difference.

Government communications must be proportionate and accurate (and we were careful that our messages were true, accurate, and avoided any deception). Therefore, in order to use an injunctive norm statement, reliable evidence of others' attitudes towards non-payment of tax was required. We were able to ensure that a survey of 1,207 UK adults which took place in May 2012 measured the extent to which respondents agreed with the statement '*Everyone in the UK should pay their tax on time*' (88% of respondents agreed).[20] We include this statement as a separate message in Experiment Two in order to operationalise the moral duty concept outlined in section 2.

The remaining messages in this group were constructed to isolate the additional effect of representing this moral duty as an injunctive norm. As outlined in section 2, the most basic

---

[20] This was an omnibus survey conducted by TNS BMRB, who kindly allowed us to include this question. A five-point Likert scale was used to measure agreement. A summary of some of the results can be found at http://www.kantar.com/public-opinion/policy/180712-attitudes-to-tax-avoidance/

formulation introduces a general injunctive norm using phrasing similar to the general descriptive norm above: "*The great majority of people agree that everyone in the UK should pay their tax on time*".[21] Rather than framing the injunctive norm as being supported by a 'great majority' of people, the exact result was presented in percentage terms ("*88% of people agree that everyone in the UK should pay their tax on time*") and as a fraction ("*Nine out of ten people agree that everyone in the UK should pay their tax on time*"). The final message in this group combined descriptive and injunctive norms: "*Nine out of ten people agree that everyone in the UK should pay their tax on time. And nine out of ten people do pay on time.*"

The third group of messages concerned two other components of the payment decision. First, we wished to investigate the effect of providing additional information about methods available for paying the tax debt. Tax researchers have increasingly argued that tax authorities should consider a more 'service oriented' approach in order to increase tax compliance (Alm & Torgler, 2011; Wenzel, 2006). Such an approach would focus on enabling compliance to be achieved more easily through the provision of information and support. We therefore included text that emphasized the choice of means to pay, and that there was an opportunity to discuss the debt: "*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.*" Interestingly, two previous studies that tested similar 'enabling' messages found that they had no significant effect on compliance (Coleman, 1996; Hasseldine et al., 2007). Second, we included a warning that interest was being charged on the debt, in order to introduce an additional salient cost to the payment decision: "*We are charging you interest on this amount.*"[22] We hypothesized that both these pieces of financial information would increase payment rates. All messages are summarized in Table 6.

6. Results for Experiment Two

As for Experiment One, we examine the effect of receiving a letter *per se* on payment rates. To do so, we create two groups – those who were randomized to be sent letters early (on August 10, 2012) and those who were sent letters late (on August 17, 2012). This is the same

---

[21] Ideally, we would have constructed messages that increase the specificity of this general injunctive norm by making the norm's source more similar to the recipient (for example, 'people in your local area agree...'). However, the survey data were not detailed enough to support such statements. As an alternative, we increased the specificity with which the level of support for the norm was presented in the messages.

[22] The letters did not state the specific interest rate, which was 3.0% during the period both experiments were conducted: http://www.hmrc.gov.uk/rates/interest-late-pay.htm

procedure as for Experiment One. When examining payment in the first nine days, we observe that 11.8% of the 21,985 from the early letter group pay. Data capture issues mean that we do not have a reliable record of early payment in the second experiment, so we use the 2.5% baseline payment rate from the first experiment. On this basis, we observe a highly significant difference between groups (9.3%, $p<0.001$).

As for the first experiment, we analyzed whether a payment had occurred in the first 23 days. To do this, we used the same logit model presented in (5), apart from the fact that we include 13 dummy variables, rather than only five. We also include two new covariates – 'accountant' and 'experienced'. The former represents whether the tax form was filed by an accountant or not, and the latter is whether the individual had been late paying his or her tax in either of the previous two tax years. We were not able to access these two variables for the first experiment. Including these two variables does not affect the treatment group coefficients, but we include them for completeness and transparency.

The trial letters significantly increased the likelihood of payment occurring in the first 23 days. As Table 8 shows, these results were maintained after data on the taxpayer's age, gender, and size of debt were added to the model. We also include whether they used an accountant or were late payers in the past. From regressions (I) and (II), it is clear that the estimates do not change. The pooled effect of the thirteen test messages was to increase payments after 23 days by 2.9 percentage points compared to the control message ($p<0.000$).

There are multiple comparisons that could be made, given the number of treatment groups we have in the second experiment. As the section above shows, the three main hypotheses in this second experiment were: the minority norm effect would be replicated; there would be a significantly different effect of descriptive and injunctive norms; and providing financial information would increase payment rates. Therefore, we focus on making these comparisons, and use a Bonferroni-adjustment while doing so. We also use the multiple hypotheses testing method proposed by List et al. (2016).

Regression (I) shows that the general injunctive norm is not particularly effective in increasing the likelihood of paying tax. The moral duty frame has a 2.2% effect, and the fraction injunctive norm has a 1.7% effect. We find that the percentage injunctive norm increases payment by 3.4% (0.07σ), which is significantly different from the fraction injunctive norm (diff=0.017, $p=0.02$). This is interesting for many reasons, since the information was the same but one was presented as a percentage and one as a fraction. One possible explanation is that greater message specificity gave the impression of greater message credibility or created a large number effect.

Finally, it is worth noting that presenting the moral duty statement as an injunctive norm did not significantly affect its impact (remember that the phrases were identical apart from the norm framing).

We next focus on the descriptive norms. The general descriptive norm increases payment by 1.4% and the local descriptive norm increases payment by 2.2%. The difference between these two effects is not significant, and we do not find support for construal level theory in these particular messages. The debt descriptive norm increases payment by 3%, and the local and debt descriptive norm has a 5.0% effect (0.11σ). This 5.0% effect is just over half the size of the 9.3% reminder letter effect mentioned above, and represents a 14.8% relative increase in payment rates. It seems that there is an additive effect of the local descriptive norm and the debt descriptive norm, since the local and debt descriptive norm is significantly different from both the local descriptive norm and the debt descriptive norm (both $p<0.01$).

Next we focus on the minority framing, which includes a minority status and a minority descriptive norm. These messages increase payments by 4.7% and 4.2%, respectively; both of these effects are significantly larger than the general descriptive norm, but are not different from the local norm effect. This result is important because it replicates the effect size from Experiment One, thus supporting one of the three main hypotheses in this second experiment. When we adjust the p-values accordingly to account for the three hypotheses, we still find that the effect of the minority descriptive norm is significant at the five per cent level ($p<0.001$).

We next analyze the second hypothesis; that is, the impact of descriptive norms versus injunctive norms. We group the treatment groups into three bundles: (i) descriptive; (ii) injunctive; and (iii) other. We compare descriptive and injunctive only, and we find that the descriptive treatment group has a 1.44% ($p<0.000$) larger effect on payment than injunctive norms. So we can argue with some precision that descriptive norms produce a larger effect on tax payments than injunctive norms.

We next analyze the two non-norm based messages. We find that telling people that they are being charged interest daily had a 3.9% effect ($p<0.001$), and providing more information on how to pay their tax increased payment by 3.2% ($p<0.001$) (both p-values corrected for multiple comparisons); the latter is significantly lower than the local and debt descriptive norm ($p=0.01$, uncorrected).

Result 6: We replicate the results of the first experiment and demonstrate the reliability of norm messages.

Result 7: Both injunctive and descriptive norm messages change taxpaying behavior, but descriptive norms have a larger impact than injunctive norms.

Result 8: Providing financial information increases tax payments.

Regression (II) includes five background variables for each individual in the dataset. The coefficients on age, gender, and initial debt all go in the same direction as Experiment One, with very similar magnitudes. It is clear that people who have an accountant are more likely to pay within the 21 days, and those who have been late in the past are also likely to be late in paying their taxes in this experiment.[23] We then examined if the effectiveness of text messages varied according to whether the recipient had previously been late with their tax (and thus will have received a reminder letter in the past). Regression (III) shows the results for those taxpayers who had been late in either of the preceding two tax years, and regression (IV) gives the results for those who were not late. The effects are clearly larger for those without a recent debt, although three of the messages still created significant (p<0.05) improvements amongst those who had been paid late recently (local and debt descriptive norm, minority status, injunctive and descriptive norm).

As for Experiment One, we present the results by quartiles of debt size (Table 9). Again, we do not observe any clear patterns by debt quartile. We also analyse the message effects over a longer period (Table 10). Regression (I) gives the results after 42 days, and Regression (II) after 70 days. What is clear across these two regressions is that the majority of message effects remain economically meaningful and statistically significant 70 days after the initial letter was issued. This result shows the persistence of these messages on people's future behavior, although we see a clear decrease in the size of the coefficients.

Regression (III) considers a different outcome variable: the impact of the messages on the number of days until payment occurs. Since this is a continuous variable, we have applied an OLS regression for this analysis. As the table shows, there a few messages that move payment dates by two days. For instance, the percentage injunctive norm makes people pay two days earlier, local and debt descriptive norm and minority status make people pay 2.8 days earlier, and the minority descriptive norm makes people pay 2.2 days earlier. It is interesting to compare these results with

---

[23] Note that these letters went directly to individual taxpayers, rather than accountants. We have not presented interactions between the treatment effects and the covariates because of the large number of treatments in this experiment.

the observable characteristics of the recipients. For instance, each extra year of age brings payment forward by 0.26 days, males pay 2.4 days later, each extra pound of initial debt postpones payment by 0.0002 days, having an accountant advances payment by 2.4 days, and those who have recently been late in paying their taxes are likely to wait 21 days longer than those who have not.

Finally, we can calculate the accelerated 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. These amounts sum to nearly £9.3 million of accelerated revenue over the 23 day period. When considering these sums, it should be noted that the costs of this intervention were very low.

Result 9: Framing led to nearly £9.3m in accelerated revenue in the first 23 days, which outperformed the first experiment.

7. Discussion

The current study adds to the field in the following ways: it allows precise, rather than estimated, measurements of non-compliance; it adds to the very limited evidence base on the problem of tax non-payment; it has sufficient power to compare the behavioral effects of a variety of messages; and it was integrated into the existing business processes of a large economy's tax collection authority. Looking at the similar field experiments surveyed by Hallsworth (2014), it appears that the sample size and external validity represent the main advantages of this study; however, other field experiments can draw on richer datasets and have the ability to test a wider range of interventions (for example, different enforcement mechanisms). Fellner et al. (2013) offers a particularly relevant comparison. Although the experimental setting has some differences – since it concerns non-payment of television licences in Austria – there are many similarities. As noted above, the effect size that Fellner et al. (2013) find for the mere receipt of a letter is very similar to the one obtained in our study. In contrast to our study, morality and norm treatments did not produce a significant result.[24]

Our results suggest that there is value in adopting an approach that incorporates moral costs into taxpayer communication. These costs can be increased by presenting messages relating to social norms (both injunctive and descriptive), public services and moral duty. In particular, we

---

[24] Similar to our results in Table 8, columns III and IV, Fellner et al. could also draw on more detailed data to show that baseline levels of compliance affected message effectiveness: in high evasion areas, there was evidence of a positive effect of norm information and a negative effect of a moral duty message.

isolate the *minority status frame* (which explicitly presents the recipient's current behavior as a minority activity) as significantly increasing the perception of benefits of conforming to the norm.[25] We also show that increasing the salience of a penalty is effective. Finally, we replicate the main findings in a second field experiment, which suggests we are not just dealing with the effects of novelty. In total, this evidence suggests that the framing of information influences tax behavior – with effects sustained 70 days after the intervention.

As noted, many other studies have found that presenting persuasive messages does not increase real-world tax compliance. Three main explanations for this discrepancy suggest themselves: (i) the messages in other trial were not effective at changing behavior, even though other messages based on the same concepts may be; (ii) decisions related to *declaring* tax are substantially different from those based on *paying* tax; (iii) other differences in institutional contexts and research designs account for the contrasting conclusions. The second explanation points towards the need for more theoretical and empirical work on tax payment rather than tax declaration. Recent field experiments are just starting to address this need.

Two limitations of these studies are worth highlighting. First, although the participants for this study were drawn from the full UK Self Assessment population in 2011 and 2012, they constitute those taxpayers who had deliberately or mistakenly failed to pay their outstanding tax debts. They may, therefore, not be representative of the taxpayer population as a whole. However, there are good reasons to think that non-compliant taxpayers are *less* likely to respond to norms messages than the population as a whole, since they may have lower tax morale than the general population. Indeed, we note that the social norm effects were stronger for the subset of our group who had not had a recent debt prior to Experiment Two. Here we can point to the contribution of Dwenger et al. (2016), who propose a conceptual framework that incorporates heterogeneity in intrinsic motivation to account for differences in responses by agents to compliance initiatives; this study also found that the offer of social recognition for tax compliance increases payments by the intrinsically motivated (baseline compliers), while reducing them for the extrinsically motivated.

Second, it is not clear the extent to which these interventions produced *new* revenue, as opposed to accelerated revenue (since there was not a 'no intervention' control group). It is worth noting that accelerating revenue brings benefits itself. In addition to suffering increased borrowing costs, governments have to employ staff to manage and collect debt, which has

---

[25] A caveat is needed here: as researchers in other fields have remarked, there are difficulties in identifying what precise aspect of a message is producing a particular effect – tone, vocabulary, length, and so on (Jackson 1992). However, we have been careful to control for these factors wherever possible (partly by keeping the messages short), and certainly to a greater extent than most previous studies in this field.

potentially large opportunity costs. These employees also have to collect any penalties or interest imposed on the debt, in addition to the debt itself. It is not the case, therefore, that late payments are good for overall welfare. We do not quantify these benefits because to do so would be extremely complex, involve multiple assumptions, and require data that have not been made public by the authorities. Instead, we focus on the data at hand (increased payments at 23 days) and are careful to present these benefits as accelerated revenue.

The results of this study have policy implications. The collection of taxation is a crucial function for governments worldwide and any interventions shown to reduce non-compliance are clearly valuable in their own right, particularly if they are very low cost to implement. Indeed, our approach has been adopted more widely in the tax authority (Her Majesty's Revenue and Customs, 2013). Future research should address the extent to which these moral concerns can influence behavior in other policy areas.

References

Akerlof, George A. & Rachel Kranton (2010) *Identity economics: How our identity shapes our work, wages, and well-being*. Princeton, NJ: Princeton University Press.

Allcott, Hunt, & Todd Rodgers (2014). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review*, 104(10), 3003-3037.

Allingham, Michael G. & Agnar Sandmo (1972) Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1, 323-338.

Alm, James (2012) Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies. *International Tax and Public Finance*, 19, 54–77.

Alm, James, Bloomquist, Kim & Michael McKee (2013) When you know your neighbour pays taxes: Information, peer effects, and tax compliance. Appalachian State University, Department of Economics Working Paper No. 13-22.

Alm, James, McClelland, Gary H. & William D. Schulze (1992) Why do people pay taxes? *Journal of Public Economics*, 48 (1), 21-38.

Alm, James & Benno Torgler (2011) Do ethics matter? Tax compliance and morality. *Journal of Business Ethics*, 101 (4), 635-651.

Alm, James, Jackson, Betty. R., & Michael McKee (1992) Estimating the determinants of taxpayer compliance with experimental data. *National Tax Journal*, 45 (1), 107-114.

Andreoni, James, Erard, Brian & Jonathan Feinstein (1998) Tax compliance. *Journal of Economic Literature*, 36, 818-860.

Ariel, Barak (2012) Deterrence and moral persuasion effects on corporate tax compliance: Findings from a randomized controlled trial. *Criminology*, 50 (1), 27-69.

Becker, Gary (1968) Crime and punishment: An economic approach. *Journal of Political Economy*, 76 (2), 169-217.

Bhargava, Saurabh & Dayanand Manoli (2012) Why are benefits left on the table? Assessing the role of information, complexity, and stigma on take-up with an IRS field experiment. University of Chicago Working Paper.

Blumenthal, Marsha, Christian, Charles & Joel Slemrod (2001) Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax Journal*, 54 (1), 125 – 138.

Boame, Attah K. (2008) A panel analysis of behavior change in individual income tax compliance. Paper presented at the 2009 IRS Research Conference, Washington, D.C. At: http://www.irs.gov/pub/irs-soi/09resconpanelan.pdf

Bobek, Donna. D., Amy M. Hageman & Charles F. Kelliher (2013) Analyzing the role of social norms in tax compliance behavior. *Journal of Business Ethics*, 115 (3), 451-468.

Bobek, Donna D. & Richard C. Hatfield (2003) An investigation of the theory of planned behavior and the role of moral obligation in tax compliance. *Behavioral Research in Accounting*, 15 (1), 13-38.

Bobek, Donna. D., Roberts, Robin W. & John T. Sweeney (2007) The social norms of tax compliance: Evidence from Australia, Singapore, and the United States. *Journal of Business Ethics*, 74 (1), 49-64.

Bott, Kristina, Alexander W. Cappelen, Erik Ø. Sørensen, Bertil Tungodden (2014) You've got mail: A randomised field experiment on tax evasion. Discussion Paper 26/2014, Department of Economics, Norwegian School of Economics.

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

Carrillo, Paul, Dina Pomeranz, Monica Singhal (2014) Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement. NBER Working Paper No. 20624.

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

Chetty, Raj & Emmanuel Saez (2009) Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients, NBER Working Paper, No. 14836.

Cialdini, Robert B. (2003) Crafting normative messages to protect the environment. *Current Directions in Psychological Science*, 12, 105-109.

Cialdini, Robert B., Kallgren, Carl A. & Raymond R. Reno (1991) A focus theory of normative conduct. *Advances in Experimental Social Psychology*, 24, 201-234.

Cialdini, Robert B. & Noah J. Goldstein (2004) Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591–622.

Clotfelter, Charles T. (1983) Tax evasion and tax rates: An analysis of individual returns. *Review of Economics and Statistics*, 65 (3), 363-373.

Coleman, Stephen (1996) The Minnesota income tax compliance experiment. Minnesota Department of Revenue, St. Paul, Minnesota.

Cowell, Frank A. & James P. F. Gordon (1988) Unwillingness to pay tax: tax evasion and public provision. *Journal of Public Economics*, 36, 305-321.

Deaton, Angus (1991) Saving and Liquidity Constraints. *Econometrica*, 59, 1221-48.

De Martino, Benedetto, Kumaran, Dharshan, Seymour, Ben & Ray J. Dolan (2006) Frames, biases, and rational decision-making in the human brain. *Science*, 313, 684–687.

Del Carpio, Lucia (2013) Are the Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru. Princeton University Working Paper.

Dwenger, Nadja, Kleven, Henrik, Rasul, Imran, and Johannes Rincke (2016) Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany. *American Economic Journal: Economic Policy* 8 (3), 203-232.

Elster, John (1989) Social norms and economic theory. *Journal of Economic Perspectives*, 3, 99–117.

Erard, Brian & Jonathan S. Feinstein (1994) The role of moral sentiments and audit perceptions in tax compliance. *Public Finance = Finance Publiques*, 49, 70-89.

Fehr, Ernst & Simon Gachter (1998) Reciprocity and economics: The economic implications of homo reciprocans. *European Economic Review*, 42 (3-5), 845-59.

Feld, Lars P. & Bruno S. Frey (2010) Tax evasion and the psychological contract. In: Alm, James, Martinez-Vazquez, Jorge & Benno Torgler (eds.) *Developing Alternative Frameworks for Explaining Tax Compliance*. pp. 74-94. Abingdon, UK: Routledge.

Feld, Lars P. & Claus Larsen (2012) *Undeclared work, deterrence and social norms: the case of Germany.* Heidelberg, Germany: Springer.

Fellner, Gerlinde, Sausgruber, Rupert & Christian Traxler (2013) Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association*, 11, 634-660.

Frey, Bruno S., & Benno Torgler (2007) Tax morale and conditional cooperation. *Journal of Comparative Economics*, 35, 136–159.

Gangl, Katharina, Torgler, Benno, Kirchler, Erich, & Eva Hofmann (2014) Effects of supervision on tax compliance: Evidence from a field experiment in Austria. *Economics Letters*, 123 (3), 378-382.

Ganzach, Yoav & Nili Karsahi (1995) Message framing and buying behavior: A field experiment. *Journal of Business Research*, 32, 11-17.

Hallsworth, Michael (2014) The use of field experiments to increase tax compliance. *Oxford Review of Economic Policy*, 30 (4), 658-679.

Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Vlaev (2015) The Making of Homo Honoratus: From Omission to Commission. NBER Working Paper, 21210.

Harrison, Glenn W. & John A. List (2004) Field Experiments. *Journal of Economic Literature*, 42 (4), 1009-1055.

Hasseldine, John (2000) Using persuasive communications to increase tax compliance: What experimental research has (and has not) told us. *Australian Tax Forum*, 15, 227-242.

Hasseldine, John & Peggy A. Hite (2003) Framing, gender and tax compliance. *Journal of Economic Psychology*, 24 (4), 517-533.

Hasseldine, John, Hite, Peggy, James, Simon & Marika Toumi (2007) Persuasive communications: Tax compliance enforcement strategies for sole proprietors. *Contemporary Accounting Research*, 24 (1), 171-194.

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

Her Majesty's Revenue and Customs (2011) *Measuring tax gaps 2011*. London: The Stationery Office.

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

Jackson, Sally (1992) *Message effects research: Principles of design and analysis.* New York: Guilford Press.

Khan, Adnan Q., Asim I. Khwaja, Benjamin A. Olken (2014). Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors. NBER Working Paper No. 20627

Kirchler, Erich (2007) *The economic psychology of tax behaviour*. Cambridge, UK: Cambridge University Press.

Kleven, Henrik J., Knudsen, Martin B., Kreiner, Claus T., Pedersen, Soren & Emmanuel Saez (2011) Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79 (3), 651–692.

Kling, Jeffrey R., Mullainathan, Sendhil, Shafir, Eldar, Vermeulen, Lee C. & Marian V. Wrobel (2011) Comparison friction: Experimental evidence from Medicare drug plans. *Quarterly Journal of Economics*, 127 (1), 199-235.

Laibson, David (1997) Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics,* 112 (2), 443-478.

Levitt, Steven D. & John A. List (2007) What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21 (2), 153-174.

Liebman, Jeffrey B. & Erzo R. Luttmer (2011) Would people behave differently if they better understood social security? Evidence from a field experiment. NBER Working Paper No. 17287.

List, John A., Azeem M. Shaikh, & Yang Xu (2016). Multiple Hypothesis Testing in Experimental Economics. NBER Working Paper No. 21875.

Maniadis, Zacharias, Tufano, Fabio, & John A. List (2014) One swallow doesn't make a summer: New evidence on anchoring effects. *The American Economic Review*, 104 (1), 277-290.

Meyerowitz, Beth E. & Shelly Chaiken (1987) The effect of message framing on breast self-examination attitudes, intentions, and behavior. *Journal of Personality and Social Psychology*, 52, 500-510.

Myles, Gareth D. & Robin A. Naylor (1996) A model of tax evasion with group conformity and social customs. *European Journal of Political Economy*, 12, 49–66.

O'Donoghue, Ted & Matthew Rabin (1999) Doing it now or later. *American Economic Review*, 89 (1), 103-124.

Phelps, Edmund S. & Robert A. Pollak (1968) On second-best national saving and game-equilibrium growth. *Review of Economic Studies*, 35 (2), 185-199.

Pomeranz, Dina (2015) No taxation without information: Deterrence and self-enforcement in the Value Added Tax. *The American Economic Review*, 105 (8), 2539-2569.

Postlewaite, Andrew (2010). Social Norms and Preferences. In Benhabib, J., A. Bisin, M. Jackson (eds.) *Handbook for Social Economics*. Vol. 1A, The Netherlands: North-Holland.

Reckers, Philip M. J., Sanders, Debra L. & Stephen J. Roark (1994) The influence of ethical attitudes on taxpayer compliance. *National Tax Journal*, 47 (4), 825-836.

Rivis, Amanda & Paschal Sheeran (2003) Descriptive norms as an additional predictor in the theory of planned behaviour: A meta-analysis. *Current Psychology*, 22 (3), 218-233.

Rogers, Todd & Erin Frey (2014) Changing behavior beyond the here and now. Harvard Kennedy School Faculty Research Working Paper RWP14-014.

Rothman, Alexander J. & Peter Salovey (1997) Shaping perceptions to motivate healthy behavior: The role of message framing. *Psychological Bulletin*, 121, 3-19.

Schultz, P. Wesley, Nolan, Jessica M., Cialdini, Robert B., Goldstein, Noah J. & Vladas Griskevicius (2007) The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18 (5), 429-434.

Slemrod, Joel (2007) Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives*, 21 (1), 25-48.

Slemrod, Joel (2016) Tax compliance and enforcement. New research and its policy implications. SSRN 2726077.

Slemrod, Joel & Caroline Weber (2012) 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 (1), 25-53.

Spicer, Michael W. & Scott B. Lundstedt (1976) Understanding tax evasion. *Public Finance= Finances publiques*, 31 (2), 295-305.

Torgler, Benno (2007) *Tax compliance and tax morale: A theoretical and empirical analysis*. Cheltenham, UK: Edward Elgar.

Torgler, Benno (2012) A field experiment on moral-suasion and tax compliance focusing on under-declaration and over-deduction. QUT School of Economics and Finance Working Paper no. 285.

Traxler, Christian (2010) Social norms and conditional cooperative taxpayers. *European Journal of Political Economy*, 26 (1), 89-103.

Trope, Yaacov & Nira Liberman (2010) Construal level theory of psychological distance. *Psychological Review*, 117, 440-463.

Tversky, Amos & Daniel Kahneman (1992) Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5 (4), 297-323.

United States Treasury (2009) Update on reducing the Federal Tax Gap and improving voluntary compliance. At: http://www.irs.gov/pub/newsroom/tax_gap_report_-final_version.pdf

Vlaev, Ivo, Chater, Nick, Stewart, Neil & Gordon D. A. Brown (2011) Does the brain calculate value? *Trends in Cognitive Sciences*, 15, 546-554.

Webley, Paul, Robben, Henry, Elffers, Henk & Frank A. Cowell (1991) *Tax evasion: An experimental approach*. New York, USA: Cambridge University Press.

Wenzel, Michael (2004) An analysis of norm processes in tax compliance. *Journal of Economic Psychology*, 25 (2), 213-218.

Wenzel, Michael (2005a) Misperception of social norms about tax compliance: From theory to intervention. *Journal of Economic Psychology*, 26 (6), 862-883.

Wenzel, Michael (2005b) Motivation or rationalisation: Casual relations between ethics, norms and tax compliance. *Journal of Economic Psychology*, 26 (4), 491-508.

Wenzel, Michael (2006) A letter from the tax office: Compliance effects of informational and interpersonal justice. *Social Justice Research*, 19 (3), 345-364.

Yitzhaki, Shlomo (1974) A note on income tax evasion: A theoretical analysis. *Journal of Public Economics*, 3 (2), 201-202.

Figure 1: Raw data on the cumulative percentage of people paying per day by treatment group for the first 23 days of the study period, Experiment One
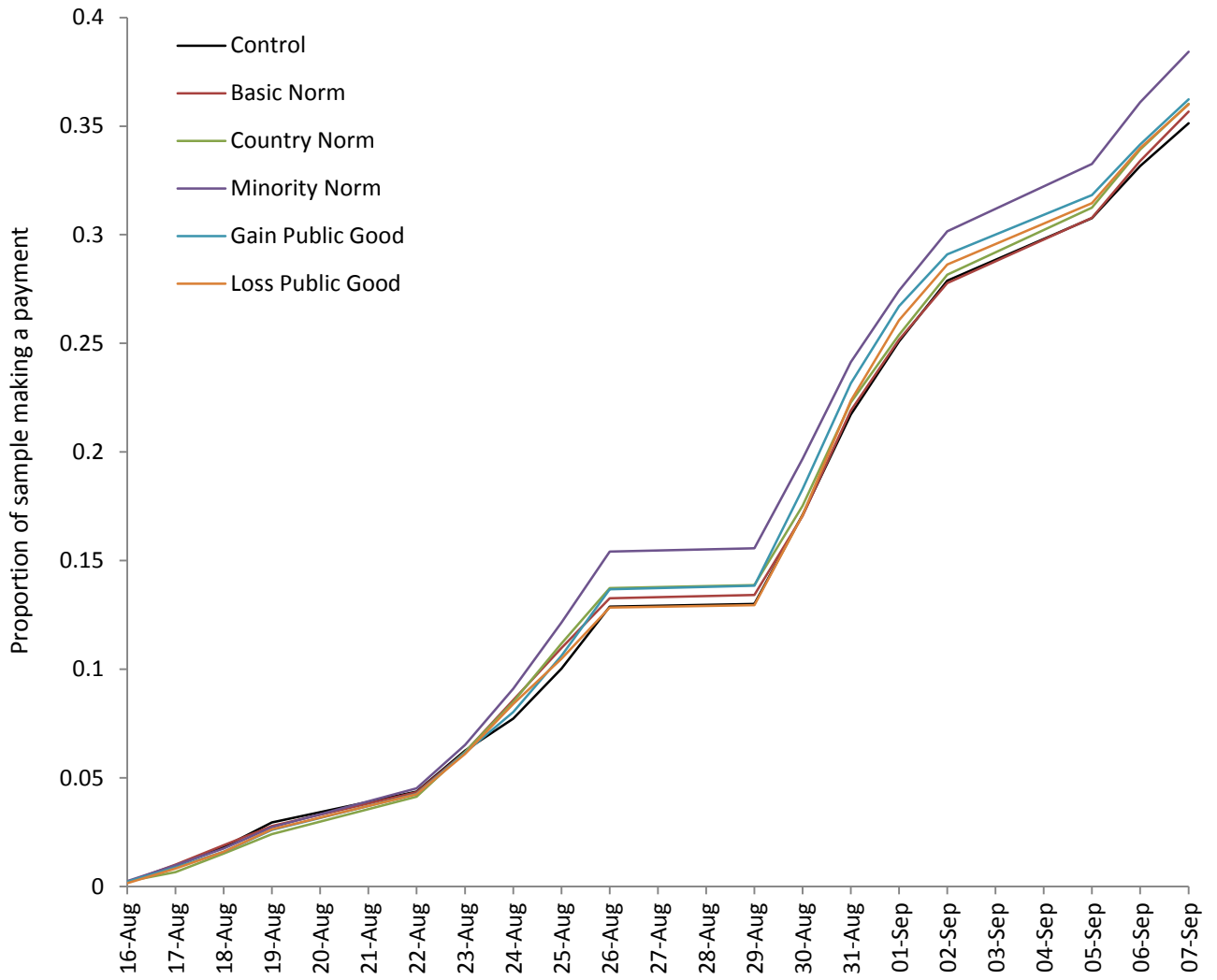
Figure 2: Raw data on the cumulative percentage of people paying per day by treatment group for the first 70 days of the study period, Experiment One
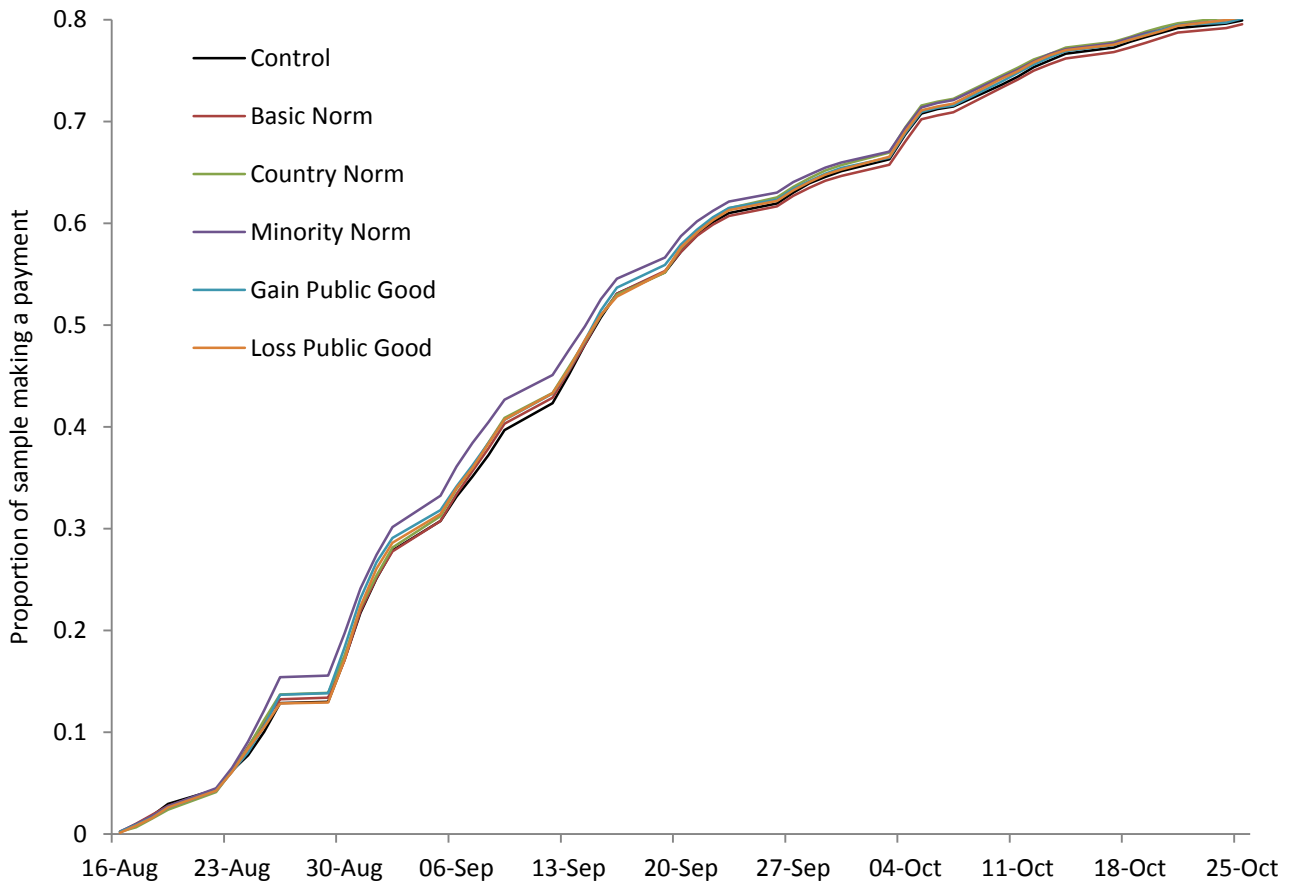
Figure 3: Raw data on the cumulative percentage of people paying per day by treatment group for the first 23 days of the study period, Experiment Two
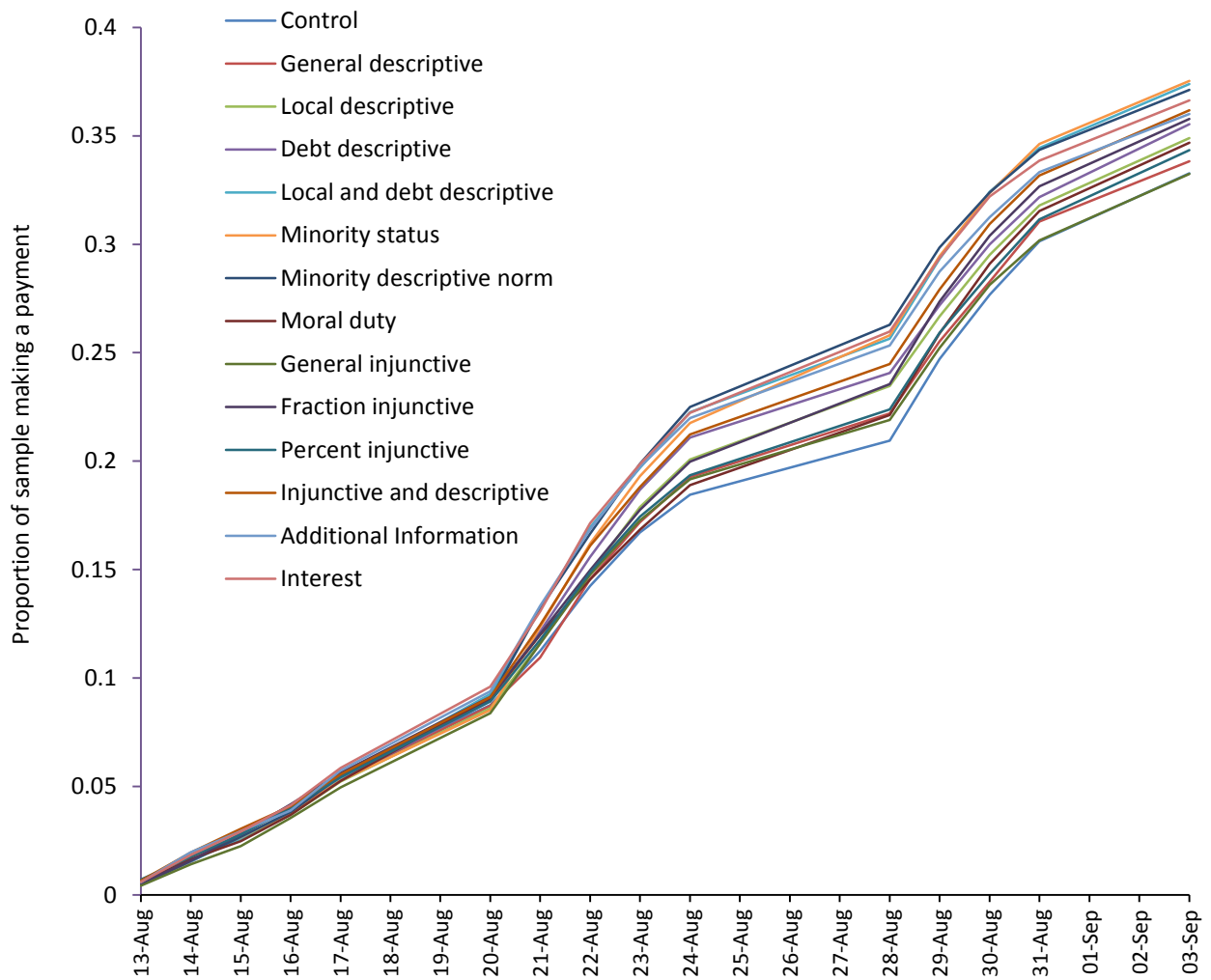
Figure 4: Raw data on the cumulative percentage of people paying per day by treatment group for the first 70 days of the study period, Experiment Two
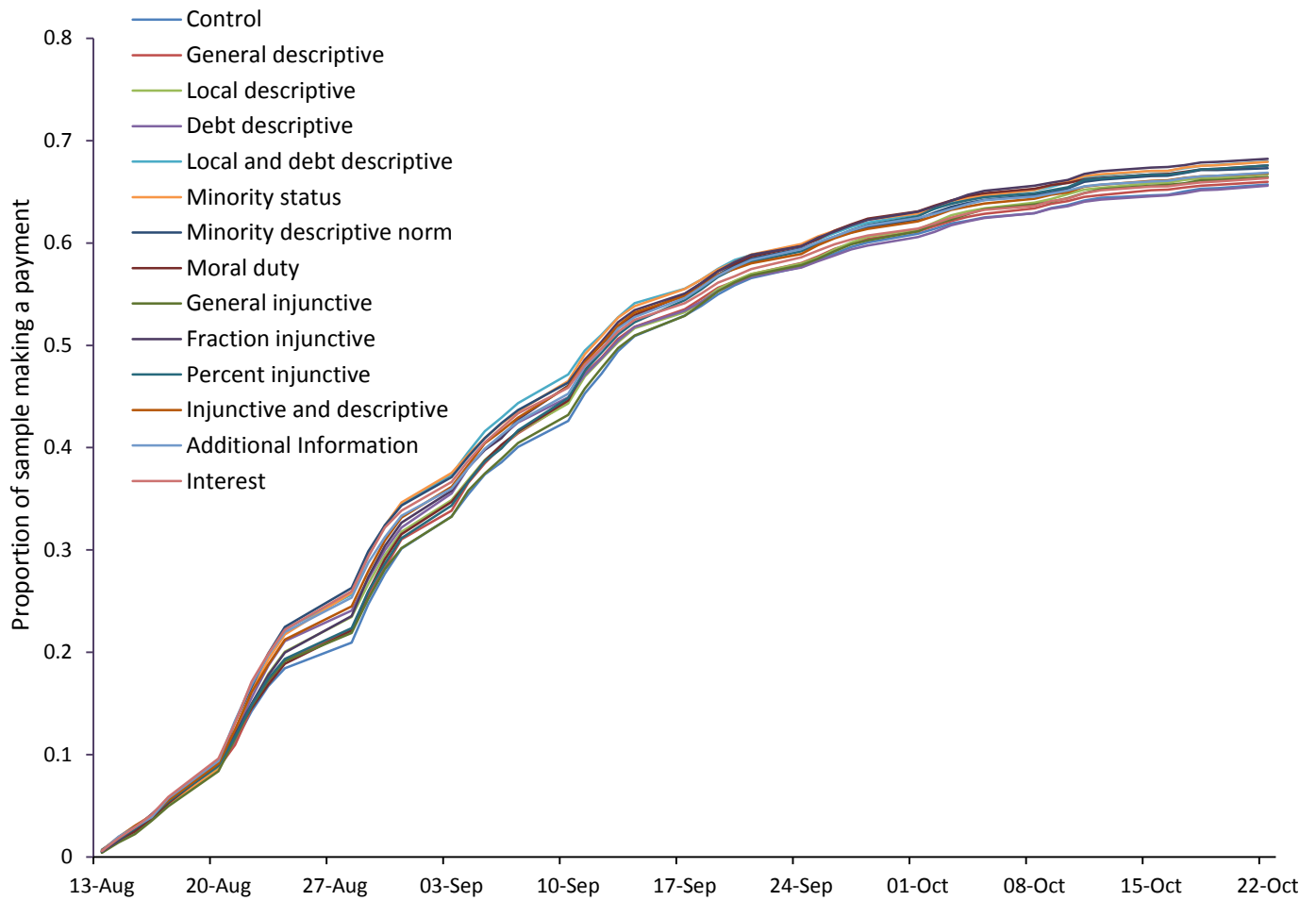
Table 1: Background characteristics of treatment groups - Experiment One

Group name	Test phrase	N	Debt value	Mean debt value	Mean Age	% Male
Control		17,038	£49,555,210	£2,908.51	49.33	73.61
Basic norm	*Nine out of ten people pay their tax on time.*	17,026	£47,923,291	£2,814.71	49.38	73.53
Country norm	*Nine out of ten people in the UK pay their tax on time.*	16,926	£46,688,514	£2,758.39	49.37	73.31
Minority norm	*Nine out of ten people in the UK pay their tax on time. You are currently in the very small minority of people who have not paid us yet.*	16,515	£46,415,638	£2,810.51	49.52	72.96
Gain-framed public services	*Paying tax means we all gain from vital public services like the NHS, roads, and schools.*	16,807	£47,640,777	£2,834.59	49.37	75.00
Loss-framed public services	*Not paying tax means we all lose out on vital public services like the NHS, roads, and schools.*	17,159	£48,875,216	£2,848.38	49.37	75.26

Table 2: Date of letter issue by treatment group - Experiment One

Group name	Day of issue					Total issue
	Tue August 16th	Wed August 17th	Thu August 18th	Fri August 19th	Mon August 22nd	
Control	4,916	3,226	2,909	3,060	2,927	17,038
Basic norm	4,695	3,277	3,022	2,990	3,042	17,026
Country norm	4,358	3,062	3,182	3,212	3,112	16,926
Minority norm	4,714	2,973	2,996	3,201	2,631	16,515
Gain-framed public services	4,611	2,922	3,154	2,926	3,194	16,807
Loss-framed public services	4,843	2,992	3,223	3,097	3,004	17,159

Table 3: Impact of the reminder letter on payment rates – Experiment One

Type of letter	% paid in first eight days	S.D.	N
Control early letter	0.092	0.290	4,916
Control late letter	0.025	0.157	2,927
Difference	**0.067*** (0.006)**		
Basic norm early letter	0.099	0.299	4,695
Basic norm late letter	0.021	0.142	3,042
Difference	**0.078*** (0.006)**		
Country norm early letter	0.095	0.293	4,358
Country norm late letter	0.024	0.153	2,112
Difference	**0.071*** (0.006)**		
Minority norm early letter	0.101	0.302	4,714
Minority norm late letter	0.024	0.153	2,631
Difference	**0.078*** (0.006)**		
Gain-public early letter	0.090	0.286	4,611
Gain-public late letter	0.031	0.173	3,194
Difference	**0.059*** (0.006)**		
Loss-public early letter	0.098	0.298	4,843
Loss-public late letter	0.022	0.148	3,004
Difference	**0.076*** (0.006)**		
All letters early letter	0.096	0.295	28,137
All letters late letter	0.025	0.155	17,910
Difference	**0.071*** (0.002)**		

Notes: Early letters were sent on August 16, 2011, and late letters were sent on August 22, 2011. Unclustered robust standard errors used for this and all subsequent tables. These results are robust to the multiple hypothesis corrections of List et al. (2016).

Table 4: Logistic regression on paying tax - Experiment One (marginal effects presented)

	(I) Pay tax	(II) Pay tax	(III) Pay tax	(IV) Pay tax after 48 days	(V) Pay tax after 70 days
Basic norm	0.013**	0.013**	0.012**	-0.005	-0.002
	(0.005)	(0.005)	(0.006)	(0.005)	(0.004)
Country norm	0.021***	0.021***	0.021***	0.009*	0.007
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Minority norm	0.038***	0.051***	0.038***	0.014***	0.010**
	(0.005)	(0.006)	(0.005)	(0.005)	(0.004)
Gain-framed	0.016***	0.016**	0.015***	0.001	0.002
public services	(0.005)	(0.006)	(0.005)	(0.005)	(0.004)
Loss-framed	0.016***	0.015**	0.015***	0.003	0.005
public services	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Age		0.005***			
		(0.0001)			
Male		-0.073***			
		(0.004)			
Debt size		2.24e-06***			
		(0.000)			
Day 2			0.003		
			(0.005)		
Day 3			0.012***		
			(0.005)		
Day 4			-0.009*		
			(0.005)		
Day 5			0.035***		
			(0.005)		
Pseudo R^2	0.00	0.01	0.01	0.01	0.01
Baseline	0.358	0.358	0.358	0.706	0.801
N	98,748	96,354	98,748	96,354	96,354

Notes: Our dependent variable is whether the individual started to pay or paid in full their outstanding tax within the 23 day period. The sample sizes are different in I vs II because data on age or gender is not available for every individual. The joint effect of the five treatments was to increase payments after 23 days by 2.2 percentage points compared to the control message (p<0.000).

Table 5: Logistic regression on paying tax - Experiment One (marginal effects presented) – split by initial debt owed

	(I) Pay tax – bottom quartile	(II) Pay tax – second quartile	(III) Pay tax – third quartile	(IV) Pay tax – top quartile	(IV) Pay tax – top 5%	(IV) Pay tax – top 1%
Basic norm	0.021*	0.016	0.002	0.011	-0.029	-0.003
	(0.011)	(0.011)	(0.011)	(0.011)	(0.024)	(0.054)
Country norm	0.032***	0.042***	0.002	0.001	0.008	0.013
	(0.011)	(0.011)	(0.011)	(0.011)	(0.025)	(0.056)
Minority norm	0.047***	0.064***	0.054***	0.036***	0.038	0.044
	(0.011)	(0.006)	(0.011)	(0.011)	(0.026)	(0.057)
Gain-framed public services	0.016	0.013	0.013	0.021*	0.018	0.084
	(0.011)	(0.011)	(0.011)	(0.011)	(0.025)	(0.054)
Loss-framed public services	0.016	0.025**	0.009	0.016	0.036	0.176***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.024)	(0.051)
Pseudo R^2	0.00	0.00	0.00	0.00	0.00	0.00
Baseline	0.378	0.329	0.346	0.378	0.424	0.393
N	24,342	24,490	24,428	24,275	4,883	967

Notes: Our dependent variable is whether the individual started to pay or paid in full their outstanding tax within the 23 day period.

Table 6: Background characteristics of treatment groups - Experiment Two

Group name	Test phrase	N	Debt value	Mean debt value	Mean Age	% Male
Control		8,558	£23,677,821	£2,766.75	50.51	71.91
General descriptive norm	The great majority of people in the UK pay their tax on time.	8,300	£22,984,178	£2,769.18	50.21	70.40
Local descriptive norm	The great majority of people in your local area pay their tax on time.	8,403	£23,592,768	£2,807.66	50.34	71.40
Debt descriptive norm	Most people with a debt like yours have paid it by now.	8,779	£24,836,091	£2,829.03	50.23	71.92
Local and debt descriptive norm	The great majority of people in your local area pay their tax on time. Most people with a debt like yours have paid it by now.	8,643	£23,563,039	£2,726.26	50.52	70.99
Minority status	You are currently in the very small minority of people who have not paid us yet.	8,587	£22,858,435	£2,661.98	50.38	70.68
Minority descriptive norm	Nine out of ten people in the UK pay their tax on time. You are currently in the very small minority of people who have not paid us yet.	8,731	£24,730,886	£2,832.54	50.44	71.72
Moral duty	Everyone in the UK should pay their tax on time.	8,507	£23,360,855	£2,746.07	50.61	71.22
General injunctive norm	The great majority of people agree that everyone in the UK should pay their tax on time.	8,595	£24,032,463	£2,796.10	50.40	71.46
Fraction injunctive norm	Nine out of ten people agree that everyone in the UK should pay their tax on time.	8,490	£22,526,004	£2,653.24	50.53	70.39
Percentage injunctive norm	88% of people agree that everyone in the UK should pay their tax on time.	8,428	£23,443,901	£2,781.67	50.47	71.18
Injunctive and descriptive norm	Nine out of ten people agree that everyone in the UK should pay their tax on time. And nine out of ten people do pay on time.	8,524	£24,175,451	£2,836.16	50.42	71.00
Additional information	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.	8,499	£23,996,925	£2,823.50	50.27	71.16
Interest	We are charging you interest on this amount.	8,483	£23,918,198	£2,819.54	50.25	70.86

Table 7: Date of letter issue by treatment group - Experiment Two

Group name	Day of issue						Total issue
	Fri August 10th 2012	Mon August 13th 2012	Tue August 14th 2012	Wed August 15th 2012	Thu August 16th 2012	Fri August 17th 2012	
Control	1,741	1,579	1,565	1,241	1,203	1,229	8,558
General descriptive norm	1,602	1,466	1,433	1,274	1,257	1,268	8,300
Local descriptive norm	1,471	1,532	1,282	1,400	1,356	1,362	8,403
Debt descriptive norm	1,566	1,477	1,607	1,470	1,299	1,360	8,779
Local and debt descriptive norm	1,584	1,464	1,447	1,505	1,270	1,373	8,643
Minority status	1,427	1,578	1,521	1,386	1,328	1,347	8,587
Minority descriptive norm	1,494	1,729	1,361	1,459	1,386	1,302	8,731
Moral duty	1,501	1,353	1,574	1,392	1,391	1,296	8,507
General injunctive norm	1,490	1,436	1,580	1,382	1,346	1,361	8,595
Fraction injunctive norm	1,458	1,431	1,512	1,352	1,335	1,402	8,490
Percentage injunctive norm	1,544	1,666	1,375	1,333	1,268	1,242	8,428
Injunctive and descriptive norm	1,847	1,327	1,273	1,261	1,546	1,270	8,524
Additional information	1,557	1,691	1,377	1,524	1,268	1,082	8,499
Interest	1,703	1,564	1,329	1,441	1,269	1,177	8,483

Table 8: Logistic regression on paying tax - Experiment Two (marginal effects presented)

	(I) Pay tax – Full sample	(II) Pay tax – Full sample	(III) Pay tax – recent debtors	(IV) Pay tax – no recent debt
General descriptive norm	0.014*	0.015*	0.004	0.026**
	0.008	0.008	0.009	0.012
Local descriptive norm	0.022***	0.023***	0.014	0.028**
	0.008	0.008	0.009	0.012
Debt descriptive norm	0.030***	0.036***	0.006	0.064***
	0.008	0.008	0.009	0.012
Local and debt descriptive norm	0.050***	0.054***	0.040***	0.064***
	0.008	0.008	0.010	0.012
Minority status	0.047***	0.052***	0.029**	0.072***
	0.008	0.008	0.009	0.012
Minority descriptive norm	0.042***	0.045***	0.014	0.074***
	0.008	0.008	0.009	0.011
Moral duty	0.022***	0.022***	0.011	0.036***
	0.008	0.008	0.009	0.012
General injunctive norm	0.006	0.005	-0.003	0.016
	0.008	0.008	0.009	0.012
Fraction injunctive norm	0.017**	0.016**	0.010	0.021*
	0.008	0.008	0.009	0.012
Percentage injunctive norm	0.034***	0.029***	0.007	0.056***
	0.008	0.008	0.009	0.012
Injunctive and descriptive norm	0.036***	0.036***	0.021**	0.050***
	0.008	0.008	0.009	0.012
Additional information	0.032***	0.035***	0.013	0.052***
	0.008	0.008	0.009	0.012
Interest	0.039***	0.040***	0.013	0.069***
	0.008	0.008	0.009	0.012
Age		0.004***		
		0.000		
Male		-0.030***		
		0.003		
Initial debt		0.00001***		
		0.000		
Accountant		0.027***		
		0.004		
Experienced		-0.219***		
		0.003		
Baseline	0.336	0.336	0.256	0.442
Pseudo R2	0.00	0.00	0.00	0.00
N	119,522	116,148	67,652	51,878

Notes: Our dependent variable is whether the individual started to pay or paid in full their outstanding tax within the 23 day period. The joint effect of the thirteen treatments was to increase payments after 23 days by 2.9 percentage points compared to the control message (p<0.000).

Table 9: Logistic regression on paying tax - Experiment Two (marginal effects presented) – split by initial debt owed

	(I) Pay tax – bottom quartile	(II) Pay tax – second quartile	(III) Pay tax – third quartile	(IV) Pay tax – top quartile	(V) Pay tax – top 5%	(VI) Pay tax – top 1%
General descriptive norm	0.024	0.020	0.005	0.007	0.032	0.014
	(0.015)	(0.015)	(0.015)	(0.015)	(0.025)	(0.034)
Local descriptive norm	0.025	0.024	0.032**	0.005	0.014	0.014

	(0.015)	(0.015)	(0.015)	(0.015)	(0.024)	(0.033)
Debt descriptive norm	0.050***	0.022	0.032**	0.016	0.025	0.005
	(0.016)	(0.015)	(0.015)	(0.015)	(0.024)	(0.033)
Local and debt descriptive norm	0.061***	0.040***	0.056***	0.046***	0.040	0.023
	(0.015)	(0.015)	(0.015)	(0.015)	(0.024)	(0.034)
Minority status	0.082***	0.042***	0.027*	0.038**	0.034	0.028
	(0.016)	(0.015)	(0.015)	(0.015)	(0.025)	(0.035)
Minority descriptive norm	0.064***	0.053***	0.028*	0.024	0.031	0.005
	(0.016)	(0.015)	(0.015)	(0.015)	(0.024)	(0.033)
Moral duty	0.037**	0.027*	0.000	0.026*	0.050**	0.013
	(0.015)	(0.015)	(0.015)	(0.015)	(0.024)	(0.034)
General injunctive norm	0.017	-0.001	-0.009	0.018	0.038	0.052
	(0.015)	(0.015)	(0.015)	(0.015)	(0.024)	(0.034)
Fraction injunctive norm	0.037**	0.011	0.004	0.016	0.016	0.006
	(0.015)	(0.015)	(0.015)	(0.015)	(0.025)	(0.033)
Percentage injunctive norm	0.048***	0.028*	0.017	0.042***	0.033	0.014
	(0.016)	(0.015)	(0.015)	(0.015)	(0.024)	(0.033)
Injunctive and descriptive norm	0.057***	0.027*	0.040***	0.018	0.029	0.006
	(0.016)	(0.015)	(0.015)	(0.015)	(0.024)	(0.033)
Additional information	0.029*	0.047***	0.031**	0.021	0.013	0.014
	(0.015)	(0.015)	(0.015)	(0.015)	(0.024)	(0.033)
Interest	0.053***	0.061*	0.008	0.036**	0.062***	0.076**
	(0.016)	(0.015)	(0.015)	(0.015)	(0.024)	(0.034)
Baseline	0.307	0.322	0.350	0.365	0.367	0.314
Pseudo R2	0.00	0.00	0.00	0.00	0.00	0.05
N	29,881	29,880	29,887	29,874	5,869	1,075

Notes: Our dependent variable is whether the individual started to pay or paid in full their outstanding tax within the 23 day period.

Table 10: Logistic and OLS regressions on paying tax and timing of tax payments - Experiment Two

	(I) Pay tax in 42 days	(II) Pay tax in 70 days	(III) Number of days late paying tax
General descriptive norm	0.009	0.003	-0.538
	(0.008)	(0.007)	0.604
Local descriptive norm	0.010	0.011	-1.136*
	(0.008)	(0.007)	0.600
Debt descriptive norm	0.013*	0.001	-0.780
	(0.008)	(0.007)	0.596
Local and debt descriptive norm	0.035***	0.024***	-2.774***
	(0.008)	(0.007)	0.595
Minority status	0.034***	0.027***	-2.808***
	(0.008)	(0.007)	0.596
Minority descriptive norm	0.023***	0.020***	-2.185***
	(0.008)	(0.007)	0.592
Moral duty	0.021***	0.019*	-1.823***
	(0.008)	(0.007)	0.595
General injunctive norm	-0.0001	0.008	-0.431
	(0.008)	(0.007)	0.595
Fraction injunctive norm	0.014*	0.018**	-1.513**
	(0.008)	(0.007)	0.596
Percentage injunctive norm	0.022***	0.023***	-1.997***
	(0.008)	(0.007)	0.595
Injunctive and descriptive norm	0.023***	0.012*	-1.540***
	(0.008)	(0.007)	0.599
Additional information	0.021***	0.014*	-1.882***
	(0.008)	(0.007)	0.598
Interest	0.015*	0.006	-1.359**
	(0.008)	(0.007)	0.600
Age			-0.257***
			0.009
Male			2.384***
			0.252
Initial debt			0.0002***
			0.000
Accountant			-2.444***
			0.309
Experienced			20.829***
			0.224
Baseline	0.509	0.644	
Pseudo R2	0.05	0.05	0.08
N	116,148	116,148	116,156

Appendix

<u>Letters</u>

Figure A1: Example of the control letter in Experiment One

Dear Sir/Madam

www.hmrc.gov.uk

Date of issue 4 August 2011
Reference REFERENCE NUMBER

Please pay £999999999999.99

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

It is easy to pay. Please call the phone number above to pay by debit card, credit card, or Direct Debit.

You can also pay using internet and telephone banking. For more information on when and 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

Figure A2: Example of the basic norm letter in Experiment One

Dear Sir/Madam

Date of issue 4 August 2011
Reference REFERENCE NUMBER

Please pay £999999999999.99

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

Nine out of ten people pay their tax on time.

It is easy to pay. Please call the phone number above to pay by debit card, credit card, or Direct Debit.

You can also pay using internet and telephone banking. For more information on when and 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

Balance checks

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

	(I)	(II)	(III)
	Male	Age	Initial debt
Basic norm	0.002	0.018	-75.737
Country norm	-0.002	0.102	-134.526**
Minority norm	-0.007	0.192	-184.245**
Gain	0.014***	0.068	-56.327
Loss	0.017***	0.052	-29.904
N	96,354	96,354	96,354

### Table A2: Regression on predictors of treatment group membership, Experiment Two

	(I)	(II)	(III)	(IV)
	Male	Accountant	Age	Initial debt
General descriptive norm	-0.014**	-0.008	-0.303	2.522
Local descriptive norm	0.000	-0.006	-0.172	40.671
Debt descriptive norm	-0.001	0.004	-0.290	62.367
Local + debt descriptive norm	-0.009	-0.009	0.006	-40.732
Minority status	-0.011	-0.014**	-0.139	-104.698
Minority descriptive norm	-0.001	-0.009	-0.081	65.548
Moral duty	-0.006	-0.002	0.093	-20.592
General injunctive norm	-0.003	-0.003	-0.114	29.434
Fraction injunctive norm	-0.014**	-0.004	0.014	-113.750*
Percentage injunctive norm	-0.005	0.000	-0.049	15.009
Injunctive and descriptive norm	-0.009	0.002	-0.094	69.173
Additional information	-0.006	0.001	-0.244	56.510
Interest frame	-0.008	-0.005	-0.264	52.555
N	116,693	119,446	119,001	119,522

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

Survival analysis

For Experiment One, we complemented our main regression specification with a survival analysis of the payment data. Figure A3 gives the Kaplan-Meier estimates for Experiment One. The dotted lines represent the three time periods in our study: the initial 23 day window; the 23-48 day period; and the 48+ day period.

Figure A3: Kaplan-Meier estimates for Experiment One
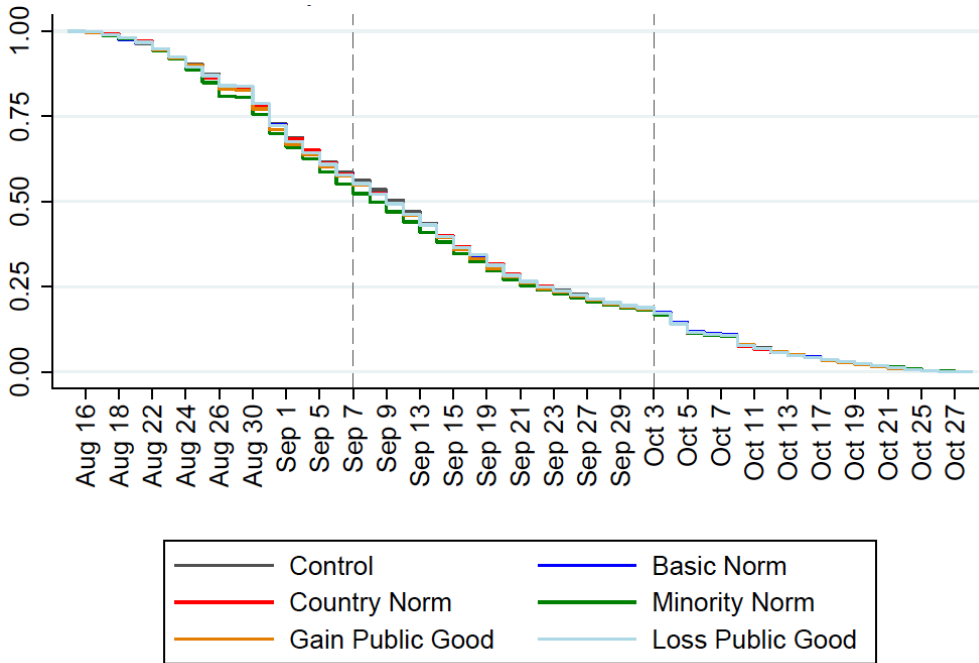
We then run log-rank tests for all treatment groups against the control group. We do this for four different time periods: for the whole study period, and then for each of the three different time periods separately (Table A3).

Table A3: Log-rank test for equality of survivor functions, Experiment One

	Entire period		0-23 days		23-48 days		48+ days	
	Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed
Control	13,648	13,813	6,417	6,330	5,156	5,032	2,075	2,081
Basic norm	13,563	13,653	6,598	6,773	4,842	4,718	2,123	2,090
Country norm	13,655	13,705	6,659	6,837	4,982	5,177	2,014	2,024
Minority norm	13,301	12,989	6,759	6,634	4,542	4,568	1,964	2,016
Gain	13,486	13,413	6,549	6,408	4,927	4,926	2,010	1,996
Loss	13,796	13,877	6,685	6,720	5,005	5,032	2,106	2,085
P value of Pr<chi2		0.033		0.001		0.005		0.715

This analysis indicates that there is a significant difference between treatment groups and control group for all time periods except the 48+ period. The survival analysis therefore produces similar results to the main regression, and also shows a decay in the treatment effects over time.

Threat perception survey

This appendix gives a brief overview of a laboratory experiment conducted to investigate the mechanisms that may be generating the taxpayer responses to the experiment. In July 2013 we recruited 231 individuals in to the Decision Research Lab at the University of Chicago. We used the downtown laboratory and experimental subject lists, which meant that we recruited mainly the general population and not students. The median age of the participants was 27 years; the proportion of participants reporting themselves to be male was 54%. We recognise that the participants were not drawn from the same country as the taxpayers in the main field experiments, and therefore may have different attitudes to tax compliance and thus different reactions to the messages.

Participants were first engaged in two rounds of experiments that focused on eliciting honest reporting of coin flips through messages that referred to acts of omission and commission; we do not present these results here. Participants were then asked three questions to establish their prior beliefs about tax compliance:

What proportion of people do you think pay their income tax on time?

In your opinion, what proportion of people think that everyone should pay their tax on time?

How acceptable do you think it is to pay your tax late?

After answering these questions, they were presented with the following text:

Imagine that you have completed an income tax form, but you have not paid the amount you owe. The deadline for payment has passed, and you have received the following letter from the tax authority.

Participants were then randomly assigned to view three of nine letters requesting payment of overdue tax owed to the Internal Revenue Service. As for the field experiment, these letters varied only by the wording of one statement (see Table A4). Given the smaller sample size, we were not able to test the full range of messages in the two field experiments. Instead, we selected seven of the messages from the field experiments, focusing mainly on descriptive norms. We also included two other messages that had featured in other field experiments, one of which is reported in Hallsworth et al. (2015). These two messages ("Omission" and "Monitoring") represent a more deterrence-based approach – both state that the government will take action in relation to the debt - and therefore offer a useful point of comparison with the norm messages. However, they are not central to the purpose of the experiment.

Table A4: Messages presented to participants in laboratory experiment

Group name	Test phrase
Country norm	*Nine out of ten people in the US pay their tax on time.*
Minority norm	*Nine out of ten people in the US pay their tax on time. You are currently in the very small minority of people who have not paid us yet.*
Minority status	*You are currently in the very small minority of people who have not paid us yet.*
Local descriptive norm	*The great majority of people in your local area pay their tax on time.*
Local and debt descriptive norm	*The great majority of people in your local area pay their tax on time. Most people with a debt like yours have paid it by now.*
Debt descriptive norm	*Most people with a debt like yours have paid it by now.*
Fraction injunctive norm	*Nine out of ten people agree that everyone in the US should pay their tax on time.*
Omission	*Previously, we treated your lack of response as an oversight. Now, if you do not contact us, we will treat this as an active choice.*
Monitoring	*We will be checking how long it takes you to respond to this letter.*

Each participant was then asked five questions. These were structured around the main determinants of the AS model (probability of detection, severity of punishment), but also included a general threat measure and a measure of message specificity:

1. How likely are you to pay the money to the government?

Please give your answer out of 10, where 0 means you are certain not to pay and 10 means you are certain to pay.

2. How certain is it that the government will catch you if you do not pay?

Please give your answer out of 10, where 0 means certainty you will not be caught and 10 means certainty that you will be caught.

3. How severe do you think the punishment would be if you are caught?

Please give your answer out of 10, where 0 means no action and 10 means the maximum penalty possible.

4. How specific is this letter to you personally?

Please give your answer out of 10, where 0 means not at all specific and 10 means very specific.

5. How threatened do you feel by this letter?

Please give your answer out of 10, where 0 means not at all threatened and 10 means very threatened.

Table A5 gives the results of an ordinary least squares regression on the participant's scores, with the Country Norm message as the comparison group.

Table A5: Results from laboratory experiment on participant responses to tax letters

	Response score				
Group name	1 – Likelihood of detection	2 – Severity of punishment	3 - Threat	4 – Specificity to recipient	5 – Likelihood of paying
Country norm	5.41	4.85	1.68	3.31	5.60
	(0.07)	(0.08)	(0.10)	(0.11)	(0.08)
Minority norm	6.22	5.73	5.18*	3.78	5.56
	(0.08)	(0.08)	(0.11)	(0.11)	(0.10)
Minority status	5.80	6.16	4.82*	2.82	5.28
	(0.08)	(0.08)	(0.11)	(0.11)	(0.08)
Local descriptive norm	6.27	6.08	5.39*	3.71	6.07
	(0.07)	(0.08)	(0.10)	(0.11)	(0.08)
Local and debt descriptive norm	5.89	6.07	5.72**	4.43	5.53
	(0.07)	(0.07)	(0.09)	(0.10)	(0.08)
Debt descriptive norm	5.85	5.57	3.31	3.07	5.33
	(0.08)	(0.09)	(0.11)	(0.13)	(0.08)
Fraction injunctive norm	6.26	5.43	3.63	3.66	6.71
	(0.07)	(0.08)	(0.11)	(0.12)	(0.08)
Monitoring	6.11	6.29	5.90**	5.39	5.47
	(0.07)	(0.08)	(0.10)	(0.13)	(0.09)
Omission	6.96[+]	6.78*	7.50***	6.71*	7.34
	(0.07)	(0.07)	(0.09)	(0.11)	(0.07)
N†	693	693	693	693	693

+ $p < 0.01$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, with individual fixed effects, clustered standard errors, and controls for gender, education level and income. The comparison group is the "Monitoring" letter.

† Since the same participant saw three letters, this N represents the number of observations. Each participant's unique identifier was retained as a control variable and the standard errors were clustered at the individual level.