

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
Imperial College London

March 2014

Abstract

Tax collection problems date back to the earliest recorded history of mankind. This paper begins with a simple theoretical construct of paying (rather than declaring) taxes, which we argue has been an overlooked aspect of tax compliance. This construct is then tested in two large natural field experiments. Using administrative data from more than 200,000 individuals in the UK, we show that including social norms and public goods messages in standard tax payment reminder letters considerably enhances tax compliance. The field experiments increased taxes collected by the Government in the sample period and were cost-free to implement, demonstrating the potential importance of such interventions in increasing tax compliance.

Keywords: taxation, natural field experiment, social norms.

JEL: C93; D03; H26

* We would like to thank Her Majesty's Revenue and Customs, the UK Cabinet Office's Behavioural Insights Team (now The Behavioural Insights Team), and TNS BMRB. In particular, Pedro Wrobel, Nick Down, David Halpern, Jackie Simms, Lesley Titley, Matt Cole, Dawn Reade, Simon Bradford, Mike Holgate, Bob Appleton, Gordon Smith, Graham Brammer, Mike Drewery, Michelle Harrison, and Ara Darzi. We thank David Novogordsky for his research support. We are grateful for the comments given by seminar participants at IFS-UCL, UCSD, Georgia State University, University of Wyoming, the 1st International Conference on Public Policy, the AEA/ASSA 2014 conference, and the 2014 FAU Taxation, Social Norms and Compliance conference; and those given by Hunt Allcott, James Alm, Stefano DellaVigna, Greer Gosnell, Erich Kirchler, Mark Phillips, Michael Price, Steven Sheffrin, and Richard Thaler. The views expressed are those of the authors, and cannot be attributed to H.M. Revenue and Customs, the Cabinet Office, or H.M. Government. Hallsworth: m.hallsworth11@imperial.ac.uk; List: jlist@uchicago.edu; Metcalfe: metcalfe@uchicago.edu; Vlaev: i.vlaev@imperial.ac.uk.

1. Introduction

Our new Constitution is now established, and has an appearance that promises permanency; but in this world nothing can be said to be certain, except death and taxes.

—Benjamin Franklin *letter to Jean-Baptiste Leroy, 1789*

Franklin's dictum on death and taxes seems undeniable today, but this was not always the case. While humans have always faced death, taxes originate from the early periods of advanced civilization. Rudimentary forms of taxation eventually evolved into property taxes, sales taxes, and inheritance taxes in the ancient states of Persia, Greece, Egypt, and Rome. Of course, with the onset of taxation comes tax evasion and avoidance (Webley et al., 1991). During the Mogul Empire (AD 1500-1750), peasants in India would abandon lands to avoid excessive taxation. Similar behavior was observed at the peak of the Roman Empire, when farmers commonly abandoned land near Rome in their flights to Constantinople to avoid land taxation. During the Middle Ages, one of the major reasons why men joined monasteries was not religious conviction, but rather a desire to avoid the ire of the tax collector.

For their part, tax collectors have developed clever schemes to collect taxes. For instance, as early as the 3rd century BC, Republican Rome sold "rent contracts" whereby the government awarded the right to collect taxes to the highest bidder in a competitive auction. This sort of "tax farming" was also used by the Ottomans, French, English, and other European governments. In medieval Egypt, Prussia, pre-revolutionary France, and China, 'share contracts' were used, whereby the government would contract with a private tax collector to collect outstanding taxes; as payment, the collector would receive a fixed percentage of the proceeds. Today we find more mundane contracting to address both evasion and avoidance: most governments simply hire their own workers to collect taxes.

Nevertheless, understanding how to motivate individuals to pay their taxes has become a major issue for economics research 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 assess why some people evade their tax payments. Their model predicts that tax evasion rates fall as the probability of detection and the degree of punishment increases. Many authors have attempted to empirically validate the AS model, with varying success; a common finding is that such models predict far too little compliance compared to observed tax behavior around the world (Alm, 1999; Torgler, 2002).

This study takes the literature in a new direction. Rather than focusing on the honest declaration of income, we use large-scale natural field experiments to learn about the factors that influence timely *payment* of taxes.¹ Such an exploration is important because an estimated 16% of the gross tax gap in the US is from enforced or late payment (Slemrod, 2007). Nevertheless, studies within the tax literature have focused primarily on tax declarations or deductions (Alm, 2012; but see Castro & Scartascini, 2013). In such studies, the dependent variable is often the amount of tax declared or deductions made, which is taken as a proxy for compliance. In the case of actual payment, no such inference is necessary: a known tax amount is outstanding, a message requests payment, and the official tax record shows whether the payment has been made. There is thus a tighter causal link between intervention and behavior. When surveying the current state of tax evasion studies, Slemrod and Weber (2012, 25-6) argue that “*the credibility revolution has, for the most part, not yet arrived, because severe measurement problems plague empirical analysis in this context.*” Our focus on payment of taxes helps to reduce these measurement difficulties.

We begin with a simple behavioral model that outlines why an agent might delay complete tax payment. The two main underlying channels for late payment include liquidity constraints and procrastination. Our definition of procrastination here is that there are people who are not liquidity-constrained who put off paying their taxes until later. While liquidity constraints are well understood among economists, reasons for procrastination can range from simply forgetting the payment is due to actively ‘putting off’ the payment because of non-pecuniary reasons (e.g. an ‘I will do it tomorrow’ strategy) (O’Donoghue & Rabin, 1999). In our model, we include moral cost (Levitt & List, 2007) as a channel that can overcome procrastination, although we do not test procrastination itself. The effect of moral cost could also be integrated into a standard cost-benefit utility framework. However, we contend that the core decision for late payers is *when* to pay, since they (unlike tax evaders) have little expectation of avoiding payment altogether. Procrastination is therefore a more appropriate framework to adopt.

We then present two natural field experiments that test the impact of tax payment reminders, focusing on carefully constructed norm and public-good messages to increase moral cost and thereby increase payment rates. This setting has several advantages. Most obviously, it provides real-world evidence concerning why people fail to pay their taxes – something we currently know little about. More generally, it provides an opportunity to test the efficacy of messaging strategies that could be applied to other policy areas. From an administrative

¹ There are only a handful of experimental studies using administrative tax data to elicit observable and precise estimates of the determinants of tax evasion – see Blumenthal et al (2001), Kleven et al. (2011), Ariel (2012), and Pomeranz (2013). Our research also speaks to the literature that suggests that information provided by governments can affect citizen behaviour (Chetty & Saez, 2009; Kling et al., 2011; Liebman and Luttmer, 2011).

perspective, identifying effective messages is an attractive strategy because they are often cost-free to introduce at scale (as was the case here).

The first natural field experiment we present was carried out with the United Kingdom tax collection authority (Her Majesty's Revenue and Customs (HMRC)) and the UK Cabinet Office's Behavioural Insights Team (BIT). It is important to note that the trials presented within this paper would not have happened without the concerted efforts of officials within the UK 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.² This paper also therefore demonstrates the value of collaborating with governments to conduct field experiments at a large scale and with a potentially large impact.

The first natural field experiment focuses on those taxpayers who have already declared their income to be taxed, but who have not yet paid their tax liabilities. Working within HMRC's existing processes, we randomized five messages across 100,000 individual taxpayers: three norm-based messages and two public goods messages. The former are: (i) "Nine out of ten people pay their tax on time" (basic norm); (ii) "Nine out of ten people in the UK pay their tax on time" (country norm); (iii) "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" (minority norm). The latter are: (iv) "Paying tax means we all gain from vital public services like the NHS, roads, and schools" (gain); (v) "Not paying tax means we all lose out on vital public services like the NHS, roads, and schools" (loss).

The experimental results show that social norm and public goods messages increase the likelihood of individuals paying their declared tax liabilities, with large differences observed within the norm messages. For example, the basic norm statement produces a treatment effect of 1.3% (0.025σ), and the country norm statement produces a treatment effect of 2.1% (0.038σ)—these effects lead to a £623,000 and £980,000 increase in total taxes paid within 23 days. In comparison, the minority norm statement produces a much larger treatment effect of 5.1% (0.1σ), which represents a £2.367 million increase in taxes paid within 23 days. In terms of gain and loss framing of the public good, we find little difference: each had an effect size of 1.6% (0.035σ), so the framing of the message had no extra impact on behavior than that from norm messaging.

² Particular acknowledgement should go to Nick Down, who read Robert Cialdini's work on social norms and made contact with the Behavioural Insights Team shortly after it was created. Michael Hallsworth implemented and managed the experiments from within HMRC. The UK BIT has also conducted experiments to understand how important text message reminders are for the payment of fines (see Haynes et al., 2013).

Overall, the results show that short messages can address non-payment through increasing moral costs. This is an important finding, since most previous studies have concluded that the framing on 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, Torgler 2004). Viewed through the lens of our model, it seems that framing messages to increase moral cost can reduce procrastination and therefore increase tax payment. In total, we estimate that more than £3 million was collected in the 23-day sample period due to the messages in the first field experiment.³ If the minority norm approach had been taken for the whole sample, £11.3 million in additional tax revenues would have been gathered by this point.

These results led to a second large-scale natural field experiment (with the same partners) that focused more carefully on message type with 119,527 UK taxpayers. This second experiment investigated whether the most effective treatment from Experiment One could be replicated. In addition, the effect of descriptive (i.e. what others do) and injunctive (i.e. what others think should be done) norms was compared. For the latter, we used injunctive messages that either stated to recipients that paying was the right thing to do or that most people thought that paying was the right thing to do. 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 our minority norm findings from the first experiment, which significantly increases the reliability of our results (Manadis et al., forthcoming). We also find large effects from the financial messages (a 3.2-3.9 percentage point increase in payment rates). Overall, we estimate that more than £9 million was collected during the sample period (i.e. 23 days) due to the messages in the second field experiment. If the most effective message had been adopted for all cases in the sample, this would have generated £15.4 million in additional revenue (in comparison to the control group). One should bear in mind that the marginal cost for the policymaker from this intervention was practically zero.

Combining results across both field experiments suggests the following conclusions. First, a model of procrastination with moral costs included is helpful for analysing why people do not pay their taxes on time. Second, while liquidity constraints have some support in our data, we find that short messages that appeal to social norms, morals and financial costs are effective at

³ The question of whether, when, at what cost the government would have received the money without these messages is extremely complicated. It is, however, clear that we cannot simply call this accelerated revenue – see footnote 19 for a fuller argument.

persuading people to pay their taxes. Third, there is great value in academics and government employees collaborating to conduct large field experiments in important policy areas – particularly those involving messages.

The remainder of the study is structured as follows. Section 2 provides a theoretical sketch that places the experiments into an economic framework. Sections 3 and 4 summarize the experimental design and results from the first natural field experiment. Sections 5 and 6 present results from the second natural field experiment. Section 7 concludes.

2. Theoretical Framework

2.1 Basic model

In the standard model of tax evasion, the taxpayer faces a decision under risk, with the extent of evasion chosen to maximize expected utility (Becker, 1968; Allingham and Sandmo, 1972; Yitzhaki, 1974). The risk arises from the possibility that the tax authority will discover the tax evasion by conducting an audit. This model has often been used to assess how much income is *declared* to tax authorities (see Alm, 2012, for a review), but it is rarely used to understand the decision to *pay* the declared income.⁴ Clearly, there are at least two stages to tax compliance. The first is to decide whether to evade. Once that decision is taken, in the second stage the individual decides to pay the declared tax on time, pay the declared tax late, or not pay the declared tax. Of course, in equilibrium the second stage reasoning affects the first stage decision, but we focus exclusively here on the second stage to provide a clear link to the natural field experiments.

In many countries, the costs of not paying declared income take the form of fines and/or interest on the outstanding tax liability. For instance, in the UK the penalty system is structured as follows. If payment is 30 days late, the agent must pay interest of five per cent on the tax that is owed at that date. If payment is six months late, the agent must pay five per cent of the tax that is owed at that date, in addition to the fine incurred at 30 days. If payment is twelve months late, the agent must pay five per cent of the tax that is owed at that date, in addition to the fines already incurred.⁵ The structure is staggered because income tax is paid in installments that are due every six months.⁶

We propose a model (with some inspiration drawn from Wang and White (2000) and

⁴ In contrast, countries such as the US and Canada explicitly refer to payment when defining compliance. See US Treasury (2009) and Boame (2008).

⁵ There is an element of O'Donoghue and Rabin's (1999) model here where agents may have present-biased preferences in determining whether to act now or wait until later.

⁶ <http://www.hmrc.gov.uk/rates/interest-late-pay.htm>; <http://www.hmrc.gov.uk/sa/deadlines-penalties.htm#6>

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

As mentioned in the prior section, we assume nonstrategic play because all of the consumers have already disclosed their income. Agents experience a disutility from having to pay their tax today ($\beta > 1$), but do not experience the same disutility from the prospect of paying the tax in the future.⁷ An agent faces a choice of whether to pay the tax now ($i = 0$) or pay the tax in the future ($i = 1$). Thus, the agent will seek to choose i to maximize the following utility function:

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

$Y = \text{realized income}$
 $M = \text{moral cost}$
 $t = \text{tax liability}$

The model begins on the self-assessment day. Our representative agent earns a stochastic labor income of Y , drawn i.i.d. from probability distribution $f(y)$. She subsequently informs the tax authority of her income which, through a deterministic rule, is converted into a tax liability which we call t , $t \geq 0$. We impose a no-borrowing condition so that our agent does not have the option of borrowing money at a prevailing market rate.⁸ Thus, if our agent experiences a negative income shock ($Y < \beta t$), luck has chosen her hand – she has no choice but $i = 1$. Otherwise, the agent has a choice of when to pay her tax. If the agent does not pay her tax immediately, interest fees accrue on her tax liability, growing geometrically at a rate of $\alpha \in (0, 1)$. On the other hand, the agent can reinvest the funds at a rate of $r \in (0, 1)$. Lastly, we assume that the agent faces a moral

⁷ This assumption is similar to the asymmetric impatience exhibited by agents with quasi-hyperbolic time preferences of the kind described in Phelps and Polak (1968), Laibson (1997), and especially O’Donoghue and Rabin (1999).

⁸ In this sense, the realized income is treated in a manner similar to “cash-on-hand” as in Deaton (1991). Explicit claims on future income cannot be used to cover the tax liability.

cost when not paying her tax liability immediately ($M > 0$), induced by a letter sent by the tax authority.

In order for her to prefer paying now (i.e., choose $i = 0$), it is sufficient that Condition 1 is met:⁹

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

Condition 1 holds as long as $\beta + r - \alpha < \frac{M}{t}$.

We introduce ϕ as the fraction of individuals that pay immediately. By defining individuals with a negative income shock in the first period ($Y < \beta t$) as liquidity-constrained, we see that the above simple model generates the following prediction:

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

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

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

These two predictions focus our attention on the effect of reminder letters and the moral costs they incur. While the use of reminder letters is quite straightforward, we will spend some time examining what we mean by moral cost in the next section.

2.2 Moral Cost

In the above model, we have a very general term for moral cost, but we can be more specific. We test two very general forms of moral cost: norms and public goods concerns. We

⁹ While we use strict preference and inequalities throughout, the results would not significantly change if we instead used weak inequalities.

propose that the way these concerns are framed can increase the effect of the M parameter on taxpayers' payment decisions. For norms, one method of framing is to make the norm more specific to the individual's situation or a group to which she belongs. Wenzel (2005) 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. For example, one can represent the average level of tax compliance for an unspecified population (e.g., "*nine out of ten people pay their tax on time*"), or one can represent the average for a specified country or population (e.g., "*nine out of ten people in [recipient's country] pay their tax on time*"). Additionally, one could also have a stronger norm that explicitly states that the message recipient is not complying with the norm, and thus is in the minority (e.g., "*nine out of ten people in [recipient's country] pay their tax on time, and you are in the small minority of people who have not yet paid their tax on time*").

These three 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. In our three framings, the psychological distance to the individual taxpayer is reduced (and hence the norm becomes more salient) as we mention his/her country of residence, and then his/her minority status in that country. Note that, strictly speaking, there is no difference in information content between letters containing messages (ii) and (iii). Both state that the taxpayer has not paid, and that nine out of ten people (i.e. a great majority) have already done so. The difference is simply whether the recipient's minority status is explicitly stated (contrast with the majority and minority frames used by Cialdini, 2003).

We also include public goods concerns, as introduced by Cowell and Gordon (1988), which allow taxpayers to derive utility from both income and public goods provision. An individual taxpayer can choose how much tax to pay but does not directly choose the quantity of public goods provided, which is determined by the tax rate and by other taxpayers' compliance decisions. We assume the taxpayer cares about income and public goods (Andreoni, 1989), but the taxpayer takes the quantity of public goods as given when choosing to pay or not. The public goods messages are framed in one of two ways: gain and loss. We assume that taxpayers are loss-

averse with respect to consumption and tax payments, so that events that have negative net utility are given disproportionate weight at the moment of choosing to pay or not (Tversky & Kahneman, 1992). We introduce the following framing using different reference points: “*Paying tax means we all gain from vital public services like the NHS [National Health Service], roads, and schools*” (gain); “*Not paying tax means we all lose out on vital public services like the NHS, roads, and schools*” (loss).

Experiment Two compares descriptive norms with injunctive norms. Cialdini et al.’s (1991) theory of normative conduct distinguishes between ‘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).

In order to use a credible 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*’.¹⁰ This statement is included as a separate treatment in Experiment Two because, although our main focus is on the impact of norms, the statement represents paying tax as a moral duty or obligation, and there have been theoretical (Erard & Feinstein, 1994), survey-based (Feld & Larsen, 2012), and experimental (Bobek & Hatfield, 2003) studies that indicate such beliefs affect tax compliance decisions.

The remaining messages in this group are constructed to isolate the additional effect of representing this moral duty as an injunctive norm. The most basic 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*’. Ideally, messages would have been included that increase the specificity of this general injunctive norm by making the norm’s context more similar to that of the recipient (for example, ‘people in your local area agree...’). However, the survey data was 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. Rather than framing the injunctive norm as being supported by a ‘great

¹⁰ 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/>

majority' of people, we presented the exact result 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.”

Importantly, we worked closely with HMRC to ensure that all these messages were accurate, credible, appropriate to the situation, and consistent with other approaches used to collect tax debts.

3. Experiment One: Methodology

3.1 Research setting

To shed light on the channels of our economic framework, we collaborated in a natural field experiment (NFE) 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 that an individual files a Self Assessment tax return, such as self-employment, multiple sources of income, or trustee status.¹¹ 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 this scheme, which 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 first 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 NFE messages into the letters sent to Self Assessment taxpayers who had not made the correct payment by July 31, 2011, and who had not responded to the initial reminder statement. All taxpayers had a debt of between £400 and £100,000 on August 1, 2011.¹² In order to minimize noise in the results, we excluded taxpayers with additional outstanding Self Assessment debts. This resulted in a sample of 101,471 individuals geographically distributed

¹¹ <http://www.hmrc.gov.uk/sa/need-tax-return.htm> - clearly this is not the same as withholding tax in the US.

¹² Debts below £400 and above £100,000 were subject to different actions and hence could not be included in the sample.

across England, Wales, and Northern Ireland. This approach and sample size permits measurement of the causal effects of norms and public goods concerns on tax paying behavior across an entire country.

3.2 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 one percentage point difference in payment rates. In addition, since the costs of running this intervention are very low (and the marginal cost of including additional cases is zero), even small effects will be cost-effective and hence worth establishing (see Duflo et al., 2008).

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. 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. 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).

We then ran regressions to establish whether membership of a treatment group was significantly predicted by any of these variables (age, gender, and size of debt). The treatment groups were largely balanced, and any differences were extremely small and statistically significant at the rate of random chance. Nevertheless, below we control for these variables in the regression specifications. We also obtain balance across time with respect to the timing of letters received ('early' versus 'late', as explained below).

Postal delivery times meant that no letters were received in the first three days after the date of issue. This pre-treatment phase allowed us to observe whether payment patterns differed between groups in the absence of any intervention. During this phase, only one group was different from the control group; the country norm group was 0.5% less likely to pay in the first three days ($p < 0.001$). The other treatment groups had no difference in their payment rates in the first three days. The lack of variation in the pre-treatment phase gives us confidence that the causal effects of any irregularities in allocation were insignificant.

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). The letters themselves did not make reference to any experimental variation in wording, and each taxpayer received a letter regardless of whether a study was taking place or not. The current experiment therefore meets the criteria of a Natural Field Experiment, as set out by Harrison and List (2004) and Levitt and List (2009).

The timeline below shows how the letters were delivered. Due to administrative policy, letters had to be delivered to all agents who had not paid their taxes. 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 the 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*. In practice, this means comparing the payment rates at August 25, 2011 of those who were issued letters on August 16, 2011 with those who were issued letters on August 22, 2011 (see timeline below).

Timeline of early vs. late letters to identify reminder effect

16 th Aug 2011 EARLY letters sent	17 th Aug 2011	18 th Aug 2011	19 th Aug 2011 EARLY letters received	22 nd Aug 2011 LATE letters sent	23 rd Aug 2011	24 th Aug 2011	25 th Aug 2011, Comparison point. LATE letters received.
--	------------------------------	------------------------------	---	---	------------------------------	------------------------------	---

3.3 Messages

All letters contained basic information on the size of debt and means of payment. The experimental variation was simply the inclusion of a short phrase, in the standard typeface, after the first sentence (see Table 1). These messages were proposed to test directly the theory outlined in section 2. See the Appendix for the control letter sent to tax payers.

These phrases were constructed to persuade the recipient to pay the amount they owe, 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.*” This phrase intends to correct any misperceptions that tax compliance is low, which could act as justification for non-payment because ‘everyone else is doing it as well’ (Wenzel 2005). Following the discussion of construal level theory above, this norm message was

constructed to be as non-specific as possible. This ‘basic norm’ is 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 a significant 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 tests 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 of the reminder letters state that the recipient has not paid her tax yet. However, as Vlaev et al. (2012) 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. To the best of our knowledge, this constitutes one of the first attempts to distinguish minority and norm effects in this way. 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). Most people realize that taxation is necessary to fund public goods that benefit everyone, as well as specific items from which they gain in particular. Thus, there are two different approaches at work. First, a simple reciprocal mechanism of ‘something for something’, wherein tax is paid in order to achieve some other kind of tangible benefit (Fehr & Gächter, 1998). Second, a more altruistic, moral perspective wherein paying tax is the means to achieve something intrinsically good (McGraw & Scholz, 1991, Reckers et al., 1994). We attempted to appeal to both of these perspectives in the fourth test phrase, which 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.*”

Of course, if paying taxes yields certain benefits, the corollary is that not paying can put them in jeopardy. Although non-payment may lead to short-term gain for an individual, since such decisions are interdependent it could also contribute to a developing ‘tragedy of the commons’ situation, where all eventually lose. Thus, it is also possible to focus on the potential loss of the benefits taxation brings. There is evidence that this different ‘goal/outcome framing’ may result in different behavioral outcomes (Meyerowitz & Chaiken, 1987; Ganzach & Karsahi, 1995; Rothman & Salovey, 1997). In order to test this framing effect, we introduced a negative [loss] frame: “*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. Our interventions are in contrast to Fellner et al. (2013), who did not use minority status for the descriptive norm, and whose moral appeal letter states that many people might not pay and that this impacts on fairness.

Each letter was addressed solely to the individual who incurred the debt, and all were sent in standard HMRC envelopes. Recipients had no way of identifying the other participants, thanks to strict taxpayer confidentiality laws, so spillover effects are not a concern.¹³ 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.

Two points relating to the messages should be noted. First, the control group is a classic control group and was not assigned a test phrase; control subjects simply receive the standard letter with basic information. Below we measure the effect of the baseline (reminder) letter itself because we staggered the date the letters were sent. However, 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.

3.4 Empirical Strategy

¹³ 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.

Our empirical strategy is to examine the impact of reminder letters and message variants on tax paying behavior. As explained in Section 3.2, to determine the impact of a reminder *per se*, we simply compare the tax payment by individuals who receive the ‘early’ letters versus individuals who receive the ‘late’ letters, using the following logit regression:

$$P_{i7} = \alpha_i + \beta_1 L_i^{early} + \varepsilon_i \quad (4)$$

where P_{i7} is whether the individual pays their tax debt in the first seven days (i.e. by August 25, 2011, which is before the individual in the ‘late’ group receives 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 message framings we run the following logit regression:

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

where P_{i23} is whether the individual pays their tax debt in the first 23 days. 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 goods treatment, and L_i^{loss} is a dummy variable for the loss-framed public goods treatment. In the results table we present marginal effects of the logit regressions.

4 Experiment One: Results

Both here and in Experiment Two, we focus on whether a payment has been made to the government. Table 3 summarizes the impact of early versus late letters to ascertain the overall effect of receiving a reminder letter as such. The first two rows in Table 3 show that the impact of the control reminder is around 6.7 percentage points - that is, those receiving a letter were nearly four times more likely to pay their tax bill than those who did not receive a letter.¹⁴ Latter 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 good framing letters also had an impact on payment, with an effect between 5.9-7.6 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

¹⁴ This is a similar effect size to Fellner et al. (2013) in the case of TV license payments.

is probably a lower bound estimate, because we could not observe whether additional payments were made after everyone received a tax letter.

Result 1: Reminder letters have an effect on tax payments.

We next turn to the impact of the messages. Figure 1 plots the percentage of people per day in the first 23 days who pay their tax back to the government in each of the six treatment groups. The days for which there is no recorded payment are weekends.¹⁵ 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.

We look more formally at these data in Table 4. These models pool all data and show how the treatments perform relative to the control across the entire sample period. Given that the test messages were only included in the first letter, 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. We have, therefore, analysed whether a payment or cleared balance had occurred by the end of the 22nd day.

We conducted a logistic regression on the individual-level data for payments and cleared balances at 22 days. The trial letters significantly increased the likelihood of payment occurring by the end of the 22nd day. As Table 4 shows, these results were maintained after data on the taxpayer's age, gender, and size of debt were added to the model. From regressions (I) and (II), it is clear that the estimates do not change apart from the minority norm treatment effect, though it is important to note that all of our estimates are statistically significant at the five per cent level as a minimum. Overall, we find that the average effect for the basic norm is 1.2% (0.025σ); while we increased the likelihood of paying debt on time, this effect is small. Similarly, we find that the effect for the country norm is 1.7% (0.035σ). The effect size for the minority norm is much larger at 4.9% above the control (0.1σ). This minority norm effect is around 70% of the reminder letter effect.

Result 2: Letters with norm statements motivate people to pay their taxes, especially when minority norm frame is used.

¹⁵ August 29th 2011 was also a public holiday in the UK, which has clearly affected the payment data.

Both the gain- and loss-framed public good messages had an effect of around 1.3% (0.027σ). Thus, we find that all of the messages have a positive impact on paying tax. We find that the minority norm has a significantly larger effect ($p < 0.01$) than any of the other four messages.

Result 3: Public goods framing matters in motivating people to pay their taxes.

Result 4: Loss-aversion framing does not influence tax payment more than gain framing.

Regressions (III) and (IV) remove the early payers (i.e. those who paid before the letters were sent out) to get a better estimate of the treatment effect. It is clear that all of the estimates increase slightly once the early payers have been removed from the sample (comparing (I) with (II)). Using regression (IV), we observe that the minority norm still produces the largest effect, amounting to 5.1% percentage points (0.11σ). 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 , $p = 0.26$). This does not support the construal level theory outlined in section 2. We interacted the letter treatments with the age, gender, and size of debt variables. None of the interactions were significant at the five percent 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).¹⁶

We also examined whether the strength of these results occurred because of a possible implicit threat that increased people's subjective value of the probability of getting caught. We ran a survey after the experiment where we randomized the same treatment letters in a lab setting and then elicited their subjective value of threat. We found that there were no differences between the control and treatment groups in terms of implicit threat of the letters.¹⁷

We can now work out the revenue received from the messages for the government at the 23-day period.¹⁸ We will use the estimates in regression (I). The minority norm message generated

¹⁶ Due to the possibility of non-linearities in the debt size variable, we performed subgroup analysis of this variable. The cases with larger tax liabilities present some interesting results. For instance, the country norm had an 8.8% (0.18σ) treatment effect for those with a £15,000 to £29,999 annual tax liability. For those with very high tax liabilities (£30,000 and above), the fairness messages both work well, with a 12% (0.24σ) and 19% (0.38σ) treatment effect for gain- and loss-framed messages respectively (the difference in loss and gain is significant, $p < 0.01$). However, the minority norm actually reduced the likelihood that people paying their tax liabilities for this specific subgroup; an 18.3% (0.36σ) reduction.

¹⁷ Results available from the authors upon request.

¹⁸ We are calling this yield rather than accelerated yield for a number of reasons. Firstly, it is not clear at what point in the debt recovery process it should be determined that a debt has not been collected. If a debt is collected 18 months later after legal action, as opposed to 3 days after a letter was issued, is this still 'accelerated revenue'? Categorizing it this way creates a crude indicator that also excludes significant benefits from reduced administrative costs and

the government £1.9 million in tax revenues that would not have been present at the 23rd day otherwise.¹⁹ If this approach had been taken on the whole sample, it would have generated the government £11.3 million in tax revenues at this point. If we take the other estimates (although not significant at the five per cent level) at face value, we can calculate the total benefits of this study. The public goods messages generated £1.5 million, while the country norm had a smaller effect, generating just £0.2 million. The basic norm actually increased the debt liability in comparison to the control by £0.4 million. In total, this amounts to £3.2 million in government revenue from taxes that would otherwise not have been paid in the experimental period.²⁰

Result 5: £3.2 million was received in the first 23 days from the behavioral framing.

Overall, the results show that there are important behavioral reasons as to why people do not pay their taxes. For example, if liquidity constraints were the sole driver of tax paying behavior, norms and public good frames should not matter. We find that they do matter, which supports further examination of the effects of norms and messaging.

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, conduct a more precise examination of what type of norm framing works, and determine whether basic financial information works. Letters were sent to Self Assessment taxpayers who had not made the correct payment by July 31, 2012, and who had not responded to the initial reminder statement. All taxpayers had a debt of between £351 and

Exchequer borrowing costs (which themselves are difficult to calculate). In practical terms, calculating new yield would introduce many more assumptions and much more complexity. It would involve introducing an arbitrary cut-off point for collected/uncollected debt (as opposed to our proposed 23-day indicator, which has a clear rationale). Also, given that the procedures used to collect debts become more tailored to debt characteristics later in the process, it would require calculating the success of every such procedure. In contrast, using increased yield after first letters is simpler and more reliable: after the first reminder letter, these payments have been made to the government that would not have been obtained if certain messages had not been used.

¹⁹ This is calculated by examining the average effect of 3.8% on the minority norm group revenue. So with 16,515 individuals in this group with a mean debt value of £2,810.51, an extra 3.8% increase in payment would be equal to £46,415,638 x 0.038.

²⁰ This leads one to think about the overall welfare costs from such interventions. It is not the case, though, that late payments are good for overall welfare. Governments have to employ thousands of staff to manage and collect debt, which has potentially large opportunity costs. These employees have to collect any penalty and interest on the debt, in addition to the debt itself.

£50,000 on August 1, 2012.²¹ 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 5).

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, and size of debt, across our fourteen groups we found that in two of those 42 instances balance was a significant factor at the 0.05 significance level. The fraction injunctive norm group was more likely to have lower mean value of debts ($p < 0.05$), and the descriptive debt norm group was more likely to have higher mean value of debt ($p < 0.05$).

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 6).

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.

²¹ 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.

The thirteen messages in the second experiment can be divided into three groups. First, six social norm messages that represent 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*”).²² 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 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 groups were intended to show 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 have therefore done so in this experiment, 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*”).

²² 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.

The second group of messages are constructed to assess the effect of messages based on injunctive norms. Cialdini et al.'s (1991) focus theory of normative conduct distinguishes between '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). 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 behaviour – 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).²³ This statement is included as a separate message in Experiment Two because, although our main focus is on the impact of norms, the statement represents paying tax as a moral duty or obligation, and there have been theoretical (Erard and Feinstein, 1994), survey-based (Feld & Larsen, 2012) and experimental (Bobek & Hatfield, 2003) studies that indicate such beliefs affect tax compliance decisions. In the terms of our discussion in Section 2, this message may directly increase the perceived moral costs of not paying.

The remaining messages in this group are constructed to isolate the additional effect of representing this moral duty as an injunctive norm. As outlined in section 2, the most basic 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*".²⁴ Rather than framing the injunctive norm as being supported by a 'great majority' of

²³ 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/>

²⁴ 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.

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; Kirchler, 2007). 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.”²⁵ We hypothesized that both these pieces of financial information would increase payment rates. All messages are summarized in Table 7.

6. Results for Experiment Two

Similar to Experiment One, our dependent variable is whether the taxpayer makes a payment. We begin by examining the effect of receiving a letter *per se* on payment rates. To do so, we create two groups – those who were randomized to receive letters early (on August 10, 2012) and those who received letters late (on August 17, 2012). When examining payment in the first eight days, we observe that 11.8% of the 21,985 from the early letter group pay within the first eight days. 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

²⁵ 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>

this basis, we observe a highly significant difference between groups (9.3%, $p < 0.001$). So we replicate the reminder effect from the first field experiment, but find a slightly larger difference.

As per the first experiment, we analyzed whether a payment or cleared balance had occurred by the end of the 22nd day. We conducted a logistic regression on the individual-level data for payments at 22 days – this was identical to the logit model presented in (5), apart from the fact that we include 13 dummy variables, rather than only five. Another small difference is that we 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 by the end of the 22nd day. As Table 7 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.

There are multiple comparisons that could be made with 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 use a Bonferroni-adjustment when making these comparisons.

We will focus on regression (I). We should recognize that the payment rate in the control group is 33.6%. We focus first on the injunctive statements. We find that the general injunctive norm is not very 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 found that the percentage injunctive norm increases payment by 3.4% (0.07σ), which is significantly different from the fraction injunctive norm ($\text{diff}=0.017$, $p=0.02$). This is interesting for many reasons, since the information was the same but one was presented in 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 together have a 5.0% effect (0.11σ). This 5.0% effect is just over half the size of the reminder letter 9.3% effect mentioned above, and represents a 14.8% relative increase in payment rates. It seems that there is additive effect of the local descriptive norm and the debt descriptive norm, since the local + 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. Both messages have powerful effects, increasing payment by 4.7% and 4.2%, respectively; both of these effects are significantly larger than the general descriptive norm, but are not different to 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 descriptive minority 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 in to 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 work better than injunctive norms for tax payment.

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 hypotheses); the latter is significantly lower than the local + 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.

We also examine the impact of whether these effects result in extra payment amounts (intensive margin) or whether the messages simply mean that people are more likely to pay altogether (extensive margin). We find no impact in terms of different payment amounts for those in the treatment groups in comparison to the control group, or for the injunctive and descriptive norm groups more generally versus the control.

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.²⁶

In regression (III) we analyze the impact of the messages in terms of how long people wait to pay. There are a few messages that mean people pay their taxes two days before the control. For instance, the percentage injunctive norm makes people pay two days earlier, local + 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. These are interesting to compare with the observable characteristics of the individual. 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 previously been late in paying their taxes are likely to wait 21 days longer than those who have never been late.

We can now turn to the overall estimates of the revenue gained from this second experiment. We can calculate the added revenue in the first 23 days by taking the coefficient for each message and multiplying it by the number of people in each treatment group, then multiplying that by the average debt. These sum to extra revenue in the first 23 days of £9.1 million.²⁷ When considering these sums, it should be noted that the costs of this intervention were virtually zero.

²⁶ Note that these letters went directly to individual taxpayers, rather than accountants.

²⁷ We therefore have the following estimates: moral duty saved £0.52 million; general injunctive norm saved £0.15 million; percentage injunctive norm saved £0.79 million; number injunctive norm saved £0.38 million; local

Result 9: Framing led to over £9.1m in extra revenue in the first 23 days, which outperformed the first experiment.

7. Discussion

7.1 Contribution to the literature

Recent years have seen growing interest in the use of field experiments to measure, explore and reduce tax non-compliance – although the number of published studies remains small. In the early part of the century, Wenzel (2005a, 2006) and Wenzel & Taylor (2004) conducted a set of artefactual field experiments on Australian businesses and individuals. These showed a significant effect of messages based on sanctions and on interpersonal fairness, but no effect of injunctive social norms. Torgler (2004 and 2012) conducted two field experiments in Switzerland that showed little or no effect of moral appeals on tax compliance.

More recently, the findings of Kleven et al. (2011)'s experiment suggest that discrepancies between the Allingham Sandmo model's predictions and observed evasion may be because third-party reporting reduces opportunities to evade. Ariel (2012) found that deterrence messages did not increase compliance amongst Israeli businesses; moral persuasion actually backfired. Pomeranz (2013) found that letters announcing an increase in audit probability increased VAT payments, but that the VAT "paper trail" had a stronger effect. This study used descriptive norms to increase tax morale but did not have the power to demonstrate their effectiveness; the norm treatment was also different to the deterrence letter in other respects, such as how to pay the tax to the government. Castro and Scartascini (2013) find that messages detailing sanctions increased compliance, but not those stressing the fairness and equity of the tax system.

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. They also find a strong effect of sending a letter *per se*, and of a threat-based message in that letter, but no effect of social norms or moral messages. Bhargav and Moli (2012) conduct a field experiment with the US IRS to increase take-up of benefits, and find significant effects of mailing a letter, simplification of information, and the display of benefits. Finally, there are many existing field experiments that deal with the effect

descriptive norm saved £0.51 million; local + debt descriptive norm saved £1.19 million; descriptive debt norm saved £0.75 million; injunctive + descriptive norm saved £0.86 million; minority status saved £1.07 million; minority norm saved £1.05 million; the interest frame saved £0.73 million; and the more information saved £0.77 million.

of social norm information in wider policy areas, most notably energy consumption (Schulz et al. 2007).

The current study adds to this field in the following ways: it allows precise, rather than estimated, measurements of non-compliance; it provides the first real-world evidence on the substantial problem of tax non-payment; it has sufficient power to make sophisticated comparisons of the behavioural effects of a variety of messages; and it is integrated into the “core business” of a large economy’s tax collection authority, thus showing the potential for collaboration between administrators, academics, and agents like the Cabinet Office Behavioural Insights Team.

This paper attempts to present the first theoretical contribution towards why people do or do not pay declared tax liabilities, and when this happens. It argues that there is value in adopt a procrastination model that incorporates moral costs as a factor that influences payment behaviour. These costs can be increased by presenting messages relating to social norms (both injunctive and descriptive), public goods and moral duties. In particular, we isolate the *minority status frame* (which explicitly links the recipient’s current behavior to a minority) as significantly increasing the perception of benefits of conforming to the norm.²⁸ We also show that increasing the salience of the penalty also makes people pay their tax on time. 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 really can change tax behavior.

This study most directly complements the best-known of the few existing tax compliance field experiments, which took place in Minnesota (Blumenthal et al., 2001). That study tested the effects of including two different state tax letters sent to taxpayers before the filing deadline. The first attempts to communicate the social norm that 93% of Minnesotans report their taxes correctly and file voluntarily. The second stresses that state tax dollars are spent on a range of public services, including education, healthcare, law enforcement, and libraries. The messages thus closely match the concepts we test in the current study. However, Blumenthal et al. find that neither of these appeals had a significant overall impact on reported income or tax liability, and they conclude that “*we find no evidence that inexpensive, mail-based appeals will significantly increase tax compliance*” (Blumenthal et al., 2001, 126). It must be noted that the Blumenthal et al. study focused on the *reporting* of liabilities, whereas we focus on *payment*.

²⁸ 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.

We do find significant increases in tax compliance, and this is not because of higher statistical power (the Minnesota study assigned almost identical numbers to its treatment arms). Three main explanations suggest themselves: (i) the messages in the Minnesota 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. Our model and two large field experiments may help to start this field of tax payment compliance.

7.2 External validity

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 (less likely, given they had already received a reminder) 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 actually *less* likely to respond to norms messages than the population as a whole. Those taxpayers with a weak intention not to pay are likely to have been removed from the sample in advance after receiving the standard reminder letter. It is likely that these taxpayers would have been more receptive to the test messages than the ones who remained in the sample. Given that these debts remained outstanding after many communications had been sent prior to the deadline - and a standard reminder afterwards - it seems reasonable to assume that a significant proportion of the remaining taxpayers were aware of the requirement to pay, but did not do so.

As mentioned above, some approaches to understanding tax compliance posit that taxpayers may possess 'tax morale' - a disposition to comply with tax law - and thus respond to normative messages. However, this approach also allows for the possibility that some taxpayers have *low* tax morale and, therefore, are less responsive to persuasive messages on the topic. The prior revealed behavior of this sample indicates that they may fall into this camp, which may increase the importance of the ensuing results. In other words, similar messages are likely to be *more* effective in the taxpayer population at large than in this sample, which points towards the first explanation for discrepancy with the Minnesota trial (that these messages were more effective).

7.3 Policy implications

The results of this study have policy implications. The collection of taxation is a crucial function for governments worldwide, and has become more so given the increased pressure on public finances since 2008. Her Majesty's Revenue and Customs estimated that the gap represented by tax that should be paid, but is not, was £35 billion in 2010 (of which 13% was estimated to be through non-payment) (Her Majesty's Revenue and Customs, 2011). In the US, the same figure is estimated to be \$290 billion (Slemrod, 2007). Therefore, any interventions shown to reduce non-compliance are clearly valuable in their own right, particularly if they are virtually free to implement. Indeed, this study generated significant revenue for the tax authority in the UK, and its approach has been adopted more widely as a result (Her Majesty's Revenue and Customs, 2013). Future research should address the extent to which these moral concerns can influence behavior in other domains.

References

- Allcott, Hunt (2011) Social norms and energy conservation. *Journal of Public Economics*, 95 (9), 1082-1095.
- Allingham, Michael G., Agnar Sandmo (1972) Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1, 323-338.
- Alm, James (1999) Tax compliance and administration. In: Hildreth, W. Bartley & James A. Richardson (eds.) *Handbook on Taxation*. New York, USA, Marcel Dekker, Inc., pp. 741-768.
- Alm, James & Benno Torgler (2011) Do ethics matter? Tax compliance and morality. *Journal of Business Ethics*, 101 (4), 635-651.
- Alm, James (2012) Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies. *International Tax and Public Finance*, 19, 54–77.
- Andreoni, James (1989) Giving with impure altruism: Applications to charity and Ricardian Equivalence. *Journal of Political Economy*, 97 (6), 1447-1458.
- 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. & 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.
- Castro, Lucio & Carlos Scartascini (2013) Tax compliance and enforcement in the Pampas: Evidence from a field experiment. International Development Bank Working Paper IDB-WP-472.

- 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.
- 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.
- Duflo, Esther, Glennerster, Rachel & Michael Kremer (2008) Using randomization in development economics research: A toolkit. In: Schultz, T. Paul & John Strauss (eds.) *Handbook of Development Economics*, Vol. 4., pp. 3895-3962.
- 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 Gächter (1998) Reciprocity and economics: The economic implications of homo reciprocans. *European Economic Review* 42 (3-5), 845-59.
- Feld, Lars P., & Claus Larsen (2012) *Undeclared work, deterrence and social norms: the case of Germany*. Heidelberg, 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.
- Ganzach, Yoav & Nili Karsahi (1995) Message framing and buying behavior: A field experiment. *Journal of Business Research*, 32, 11-17.
- Gross, Tal, Notowidigdo, Matthew J. & Jialan Wang (2013) Liquidity constraints and consumer bankruptcy: Evidence from tax rebates. *Review of Economics and Statistics*, forthcoming.

- 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.
- Haynes, Laura C., Green, Donald P., Gallagher, Rory, John, Peter & David J. Torgerson (2013) Collection of delinquent fines: An adaptive randomized trial to assess the effectiveness of alternative text messages. *Journal of Policy Analysis and Management*, 32, 718-730.
- 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.
- 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.
- Levitt, Steven D. & John A. List (2009) Field experiments in economics: The past, the present and the future. *European Economic Review*, 53, 1-18.

- 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.
- Maniadis, Zacharias, Tufano, Fabiano & John A. List (forthcoming) One swallow doesn't make a summer: how economists (mis)use experimental methods and their results. *American Economic Review*, forthcoming.
- McGraw, Kathleen M. & John T. Scholz (1991) Appeals to civic virtue versus attention to self-interest: Effects on tax compliance. *Law and Society Review*, 25, 471-498.
- 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.
- O'Donoghue, Ted & Matthew Rabin (1999) Doing it now or later. *American Economic Review*, 89 (1), 103-124.
- O'Keefe, Daniel J. (2012) From psychological theory to message design: Lessons from the story of gain-framed and loss-framed persuasive appeals. In Cho, Hyunyi (ed.) *Health communication message design: Theory, research, and practice*. Thousand Oaks, CA: Sage Publications, pp. 3-20.
- Phelps, Edmund S. & Robert A. Pollack (1968) On second-best national saving and game-equilibrium growth. *Review of Economic Studies*, 35 (2), 185-199.
- Pomeranz, Dina (2013) No taxation without information: Deterrence and self-enforcement in the Value Added Tax. Harvard Business School Working Paper.
- 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.
- 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. & Vlasdis 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 & 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.

- Torgler, Benno (2002) Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. *Journal of Economic Surveys*, 16 (5), 657-683.
- Torgler, Benno (2002) Moral-suasion: An alternative tax policy strategy? Evidence from a controlled field experiment in Switzerland. *Economics of Governance* 5 (3), 235-253.
- 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.
- 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.
- Wang, Hung-Jen & Michelle J. White (2000) An optimal personal bankruptcy procedure and proposed reforms. *Journal of Legal Studies*, 29, 255-286.
- 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.
- Wenzel, Michael & Natalie Taylor (2004) An experimental evaluation of tax-reporting schedules: a case of evidence-based tax administration. *Journal of Public Economics*, 88 (12), 2785-2799.
- Yitzhaki, Shlomo (1974) A note on income tax evasion: A theoretical analysis. *Journal of Public Economics*, 3 (2), 201-202.

Other work using field experiments that might be of interest to readers:

- List, J. A. 2004. "Young, selfish and male: field evidence of social preferences." *Econ. J.* 114, 121–149.
- List, John A. "The Market for Charitable Giving," *Journal of Economic Perspectives*, (2011), 25(2): pp. 157-180.
- List, John A. and Charles F. Mason, "Are CEOs Expected Utility Maximizers?" *Journal of Econometrics*, (2011), 162 (1): pp. 114-123.
- List, John A., "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions," *Journal of Political Economy*, (2006), 114(1), pp. 1-37.
- List, John A., "Friend or Foe? A Natural Experiment of the Prisoner's Dilemma," *Review of Economics and Statistics*, (2006), 88(3), pp. 463-471.
- List, John A. and Yana Peysakhovich, "Charitable Donations are More Responsive to tock Market Booms than Busts," *Economics Letters*, (2011), 110 (2): pp. 166-169.
- List, John A., "Social Preferences: Some Thoughts from the Field," *Annual Review of Economics* (2009), 1, pp. 563-579.
- List, John A. and Michael K. Price, "The Role of Social Connections in Charitable Fundraising: Evidence from a Natural Field Experiment," *Journal of Economic Behavior and Organization*, (2009), 69(2), pp. 160-169.
- List, John A. "The IRB Is Key in Field Experiments," *Science*, (2009), 323(5915), pp. 713-714.
- List, John A., "Introduction to field experiments in economics," *Journal of Economic Behavior and Organization*, (2009), 70(3), pp. 439-442.
- List, John A., "Informed Consent in Social Science," *Science*, (2008), 322(5902), p. 672.
- List, John A. and Daniel L. Millimet, "The Market: Catalyst for Rationality and Filter of Irrationality," *The B.E. Journal of Economic Analysis & Policy*, (2008), 8(1 - Frontiers), Article 47.
- List, John A. and David Reiley, "Field Experiments," *The New Palgrave Dictionary of Economics*, Steven N. Durlauf and Lawrence E. Blume, eds., Palgrave Macmillan Publishing, (2008).
- List, John A., "Homo experimentalis Evolves," *Science*, (2008), 321(5886), pp. 207-208.
- List, John A. and Todd Cherry, "Examining the role of fairness in high stakes allocation decisions," *Journal of Economic Behavior and Organization*, (2008), 65(1), pp. 1-8.
- List, John A., "On the Interpretation of Giving in Dictator Games," *Journal of Political Economy*, (2007), 115(3), pp. 482-494.
- List, John A. "Field Experiments: A Bridge between Lab and Naturally Occurring Data," *The B.E. Journal of Economic Analysis & Policy*, (2006), 6(2 - Advances), Article 8.
- List, John A. and Daniel M. Sturm. "How Elections Matter: Theory and Evidence from Environmental Policy," *Quarterly Journal of Economics*, (2006), 121(4), pp. 1249-1281.
- List, John A., "Using Hicksian Surplus Measures to Examine Consistency of Individual Preferences: Evidence from a Field Experiment," *Scandinavian Journal of Economics*, (2006), 108(1), pp. 115-134.
- List, John A., "Scientific Numerology, Preference Anomalies, and Environmental Policymaking," *Environmental and Resource Economics*, (2005), 32(1), pp. 35-53.
- List, John A. and Michael K. Price, "Conspiracies and Secret Price Discounts in the Marketplace: Evidence from Field Experiments," *Rand Journal of Economics*, (2005), 36(3), pp. 700-717.
- List, John A., "Testing Neoclassical Competitive Theory in Multilateral Decentralized Markets," *Journal of Political Economy*, (2004), 112(5), pp. 1131-1156.17

- List, John A., Robert P. Berrens, Alok K. Bohara, and Joe Kerkvliet, "Examining the Role of Social Isolation on Stated Preferences," *American Economic Review*, (2004), 94(3), pp. 741-752.
- List, John A., "Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace," *Econometrica*, (2004), 72(2), pp. 615-625.
- List, John A., "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field," *Quarterly Journal of Economics*, (2004), 119(1), pp. 49-89
- List, John A., "Young, Selfish, and Male: Field Evidence of Social Preferences," *Economic Journal*, (2004), 114(492), pp. 121-149.
- List, John A., McHone, W. Warren, and Daniel L. Millimet, "Effects of environmental regulation on foreign and domestic plant births: is there a home field advantage?" *Journal of Urban Economics*, (2004), 56(2), pp. 303-326.
- List, John A., Daniel L. Millimet, and W. Warren McHone, "The Unintended Disincentive in the Clean Air Act," *Advances in Economic Analysis & Policy*, (2004), 4(2), Article 2.
- List, John A., "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, (2003), 118(1), pp. 41-71.
- List, John A., Daniel L. Millimet, Per G. Fredriksson, and W. Warren McHone, "Effects of Environmental Regulations on Manufacturing Plant Births: Evidence from a Propensity Score Matching Estimator," *Review of Economics and Statistics*, (2003), 85(4), pp. 944-952.
- List, John A., W. Warren McHone, and Daniel Millimet, "Effects of air quality regulation on the destination choice of relocating plants," *Oxford Economic Papers*, (2003), 55(4), pp. 657-678.
- List, John A., "Using Random n th Price Auctions to Value Non-Market Goods and Services," *Journal of Regulatory Economics*, (2003), 23(2), pp. 193-205.
- List, John A. and Daniel Rondeau, "The impact of challenge gifts on charitable giving: an experimental investigation," *Economics Letters*, (2003), 79(2), pp. 153-159.
- List, John A., "Preference Reversals of a Different Kind: The 'More is Less' Phenomenon," *American Economic Review*, (2002), 92(5), pp. 1636-1643.
- List, John A. and David Lucking-Reiley "The Effects of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a University Capital Campaign." *Journal of Political Economy*, February 2002, 110(8), pp. 215-233.
- List, John A., "Testing neoclassical competitive market theory in the field," *Proceedings of the National Academy of Science*, (2002), 99(24), pp. 15827-15830.
- List, John A., "Preference Reversals of a Different Kind: The 'More is Less' Phenomenon," *American Economic Review*, (2002), 92(5), pp. 1636-1643.
- List, John A. and David Lucking-Reiley "The Effects of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a University Capital Campaign." *Journal of Political Economy*, February 2002, 110(8), pp. 215-233.
- List, John A., "Testing neoclassical competitive market theory in the field," *Proceedings of the National Academy of Science*, (2002), 99(24), pp. 15827-15830.
- List, John A., Erwin H. Bulte, and Jason F. Shogren, "Beggars thy neighbor: Testing for free riding in state-level endangered species expenditures," *Public Choice*, (2002), 111(3-4), pp. 303-315.
- List, John A. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sports cards," *American Economic Review*, (2001), 91(5), pp. 1498-1507. 18

- List, John A. and Charles F. Mason, "Optimal Institutional Arrangements for Transboundary Pollutants in a Second-Best World: Evidence from a Differential Game with Asymmetric Players," *Journal of Environmental Economics and Management*, (2001), 42(3), pp. 277-296.
- List, John A. and Craig A. Gallet, "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? Evidence from a Meta-Analysis," *Environmental and Resource Economics*, (2001), 20(3), pp. 241-254.
- List, John A. and David Lucking-Reiley, "Demand Reduction in a Multiunit Auctions: Evidence from a Sports Card Field Experiment," *American Economic Review*, (2000), 90(4), pp. 961-972.
- List, John A. and Shelby Gerking, "Regulatory Federalism and Environmental Protection in the United States," *Journal of Regional Science*, (2000), 40(3), pp. 453-471.
- List, John A. and Catherine Y. Co, "The Effects of Environmental Regulations on Foreign Direct Investment," *Journal of Environmental Economics and Management*, (2000), 40(1), pp. 1-20.
- List, John A. and Mitch Kuncze, "Environmental Protection and Economic Growth: What Do the Residuals Tell Us?" *Land Economics*, (2000), 76(2), pp. 267-282.
- List, John A. and Todd L. Cherry, "Learning to Accept in Ultimatum Games: Evidence from an Experimental Design that Generates Low Offers," *Experimental Economics*, (2000), 3(1), pp. 11-29.
- List, John A. and Jason F. Shogren, "The Deadweight Loss from Christmas: Comment," *American Economic Review*, (1998), 88(5), pp. 1350-1355.
- List, John A. and Jason F. Shogren, "Calibration of the difference between actual and hypothetical valuations in a field experiment," *Journal of Economic Behavior and Organization*, (1998), 37(2), pp. 193-205.
- List, John A., Michael Margolis, and Jason F. Shogren, "Hypothetical-actual bid calibration of a multigood auction," *Economics Letters*, (1998), 60(3), pp. 263-268.

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

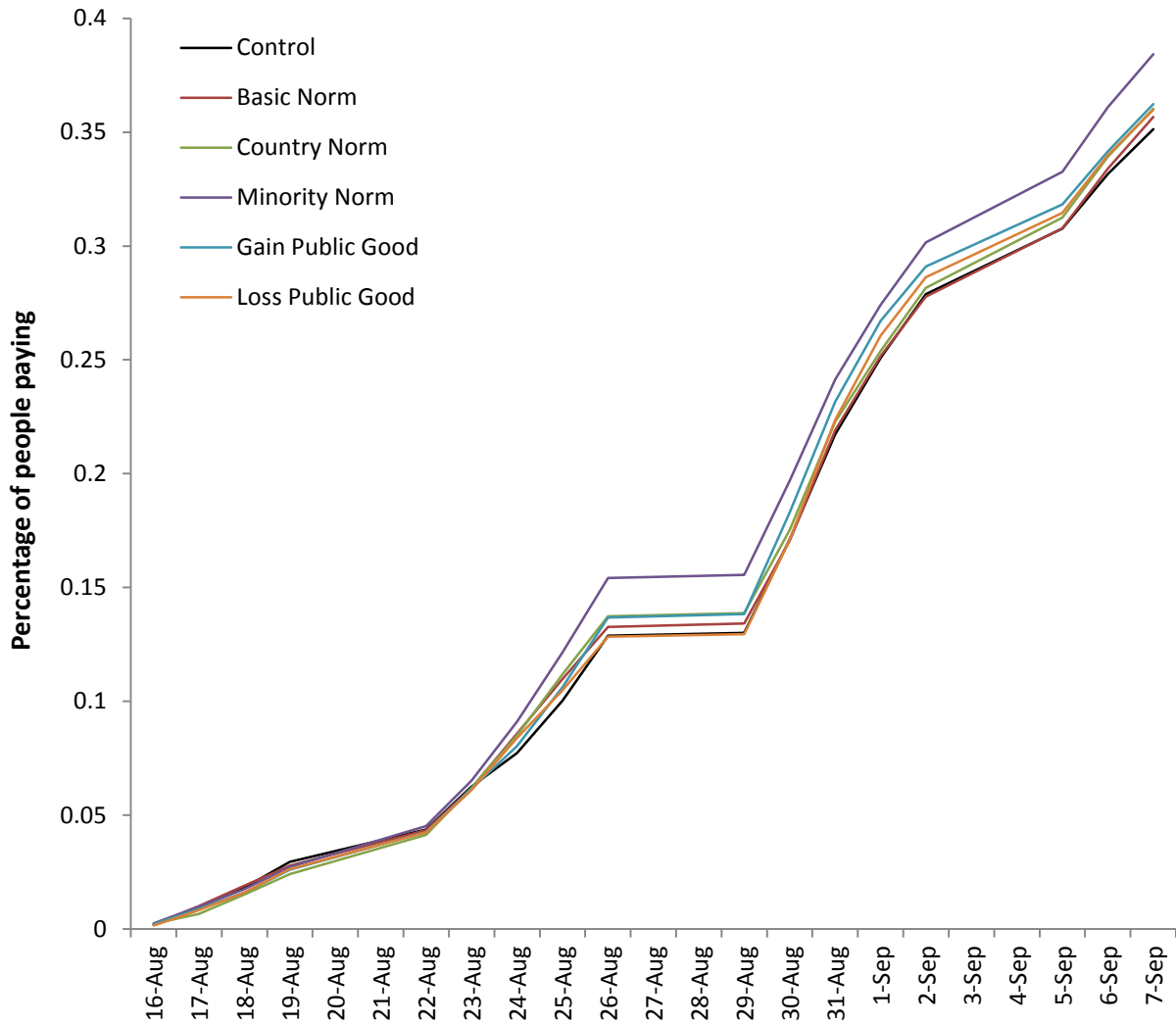


Table 1: Background characteristics of the six different groups - Experiment One

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	<i>Nine out of ten people pay their tax on time.</i>	17,026	£47,923,291	£2,814.71	49.38	73.53
Country norm	<i>Nine out of ten people in the UK pay their tax on time.</i>	16,926	£46,688,514	£2,758.39	49.37	73.31
Minority norm	<i>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.</i>	16,515	£46,415,638	£2,810.51	49.52	72.96
Gain-framed public good	<i>Paying tax means we all gain from vital public services like the NHS, roads, and schools.</i>	16,807	£47,640,777	£2,834.59	49.37	75.00
Loss-framed public good	<i>Not paying tax means we all lose out on vital public services like the NHS, roads, and schools.</i>	17,159	£48,875,216	£2,848.38	49.37	75.26

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

Group name	Day of issue					Total issue
	Tue August 16th	Wed August 17 th	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 good	4,611	2,922	3,154	2,926	3,194	16,807
Loss-framed public good	4,843	2,992	3,223	3,097	3,004	17,159

Table 3: The impact of the reminder letter on payment – Experiment One

Type of letter	% paid in first seven 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 norms early letter	0.099	0.299	4,695
Basic norms late letter	0.021	0.142	3,042
Difference	0.078*** (0.006)		
Country norms early letter	0.095	0.293	4,358
Country norms late letter	0.024	0.153	2,112
Difference	0.071*** (0.006)		
Minority norms early letter	0.101	0.302	4,714
Minority norms late letter	0.024	0.153	2,631
Difference	0.078*** (0.000)		
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.

Table 4: Logistic regression on paying some of the balance - Experiment One

	(I) Pay tax	(II) Pay tax	(III) Pay tax	(IV) Pay tax
Basic norm	0.011** (0.005)	0.012** (0.005)	0.013** (0.005)	0.013** (0.005)
Country norm	0.017*** (0.005)	0.017*** (0.005)	0.021*** (0.005)	0.021*** (0.005)
Minority norm	0.035*** (0.005)	0.049*** (0.006)	0.038*** (0.005)	0.051*** (0.006)
Gain-framed public good	0.013** (0.005)	0.013** (0.005)	0.016*** (0.005)	0.016** (0.006)
Loss-framed public good	0.013** (0.005)	0.012** (0.005)	0.016*** (0.005)	0.015** (0.005)
Age		0.005*** (0.0001)		0.005*** (0.0001)
Male		-0.073*** (0.004)		-0.073*** (0.004)
Debt size		2.37e-06*** (0.000)		2.24e-06*** (0.000)
Remove early payers	No	No	Yes	Yes
N	101,471	99,033	98,748	96,354
Pseudo R ²	0.00	0.01	0.00	0.01

Notes: Our dependent variable is whether they started to pay or paid in full their outstanding tax within the 23 day period. The sample sizes are different in I vs II, and III vs IV because not everyone has data on age or gender.

Table 5: Background characteristics of the different 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
Number 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 6: Date of letter issue by group - Experiment Two

Group name	Day of issue						Total issue
	Fri August 10 th 2012	Mon August 13 th 2012	Tue August 14 th 2012	Wed August 15 th 2012	Thu August 16 th 2012	Fri August 17 th 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
Number 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 7: Logistic regression on paying tax - Experiment Two

	(I) Pay tax	(II) Pay tax	(III) Number of days late paying tax
General descriptive norm	0.014* 0.008	0.015* 0.008	-0.538 0.604
Local descriptive norm	0.022*** 0.008	0.023*** 0.008	-1.136* 0.600
Debt descriptive norm	0.030*** 0.008	0.036*** 0.008	-0.780 0.596
Local and debt descriptive norm	0.050*** 0.008	0.054*** 0.008	-2.774*** 0.595
Minority status	0.047*** 0.008	0.052*** 0.008	-2.808*** 0.596
Minority descriptive norm	0.042*** 0.008	0.045*** 0.008	-2.185*** 0.592
Moral duty	0.022*** 0.008	0.022*** 0.008	-1.823*** 0.595
General injunctive norm	0.006 0.008	0.005 0.008	-0.431 0.595
Number injunctive norm	0.017** 0.008	0.016** 0.008	-1.513** 0.596
Percentage injunctive norm	0.034*** 0.008	0.029*** 0.008	-1.997*** 0.595
Injunctive and descriptive norm	0.036*** 0.008	0.036*** 0.008	-1.540*** 0.599
Additional information	0.032*** 0.008	0.035*** 0.008	-1.882*** 0.598
Interest	0.039*** 0.008	0.040*** 0.008	-1.359** 0.600
Age		0.004*** 0.000	-0.257*** 0.009
Male		-0.030*** 0.003	2.384*** 0.252
Initial debt		0.00001*** 0.000	0.0002*** 0.000
Accountant		0.027*** 0.004	-2.444*** 0.309
Experienced		-0.219*** 0.003	20.829*** 0.224
Pseudo R2	0.00	0.00	0.08
N	119,522	116,148	116,156

Notes: Our dependent variable for (I) and (II) is whether the taxpayer started to pay or paid in full their outstanding tax within the 23 day period. The sample sizes are different in I vs II, III because not everyone has data on age or gender.

Appendix: An example of the control letter in the first experiment

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

IDMS89P

HMRC 10/10

An example of the basic norm letter in the first experiment

Dear Sir/Madam

www.hmrc.gov.uk

Date of issue 4 August 2011

Reference REFERENCE NUMBER

Please pay £99999999999.99

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

IDMS88P

HMRC 10/10