Experimental Evidence on Taxpayer Compliance: Evidence from Students and Taxpayers

Miguel A. Fonseca
University of Exeter

Gareth Myles
University of Exeter and Institute for Fiscal Studies

Research Report 198

Acknowledgments

The authors would like to thank Shelley Hodgson, Maya Middlemiss and Colleen Saddler at Saros Research for their help in recruiting the non-student sample. The authors would also like to thank Tim Miller for programming the experimental software and Joanne Laban for conducting part of the sessions.

Disclaimer

The views in this report are the authors' own and do not necessarily reflect those of HM Revenue & Customs

Table of Contents

Executive Summary

- The most recent HMRC estimate of the tax gap for Income Tax, National Insurance Contributions, and Capital Gains Tax was £14.4 billion in 2010-11, which is about 45% of the overall tax gap (HMRC, 2012). In order to reduce this gap, it is important to understand the processes at work when individuals make a compliance decision.

- The analysis of data derived from the outcome of audits only provides information on the compliance behaviour of those taxpayers who were audited, as this is the only way of finding out for certain how much income an individual has failed to declare. Consequently, the behaviour of non-audited taxpayers has to be inferred from that of those audited. Experiments allow this information barrier to be overcome and provide enhanced understanding of taxpayer compliance.

- A laboratory experiment provides control over how different factors such as the audit rate are varied and can accurately measure the effect of each factor on compliance. This is different from the real world, or the 'field', where such control cannot be achieved.

- In addition, the variation can be conducted systematically and independently from other factors such as economic cycles. This, added to more accurate measurement of behavioural responses, makes experiments an important aid to the understanding of the relationship between operational measures and compliance.

- This report provides the first behavioural study of tax evasion and its determinants using UK taxpayers. The fundamental goal of the study is to understand what role social networks play in disseminating information about audit rates and non-compliance penalties among taxpayers. The study also determines how responsive taxpayers are to variations in audit rates and financial penalties for evasion.

- We report on data from an experiment that took place between January and March 2012. The experiment involved 1,023 participants, of whom 500 were undergraduate students at the University of Exeter (the typical participant pool in economic experiments), and 523 UK non-student taxpayers. The sessions took place online and at the FEELE Lab (the dedicated facility at the University of Exeter for conducting economic experiments).

- In the experiment participants earned income by performing a task. They then had to declare their earnings from the task to the (experimental) tax authority, which audited their tax returned randomly. If audited and caught under-reporting their earnings, participants were made to pay a fine in addition to paying the taxes on their undeclared income.

- The main treatments in the experiment manipulated:
    - The degree of knowledge about the audit rate and the rate of fine.
    - When either the audit rate and/or the fine rate were unknown, then the ability of participants to communicate as well as the scope of communication (e.g. how many people each participant could communicate with) was manipulated.

    - A treatment where the tax authority made public announcements about the number of taxpayers caught under-reporting was also implemented.

- A central finding of the experiment is that student subjects behaved in a significantly different way to non-student (or 'worker') subjects.
    - On average, students evaded more frequently than workers, and declared less income when they chose to evade.
    - As such, when there was perfect information about fine levels and audit rates, students were much more responsive to changes in fine levels than workers.
    - Surprisingly, students were particularly responsive to increases in fine levels, and unresponsive to higher likelihood of being audited.
    - The overall compliance rate of workers was unresponsive to either the fine rate or the audit rate.
    - However, there was some evidence that workers were less likely to choose to evade when both the audit and fine rates were high.

- The difference in behaviour between the two participant pools is mainly driven by the extremely high average compliance rate of 84% for workers compared to 60% for students. This is consistent with existing evidence that taxpayers are attitudinally compliant (Barham and Fox, 2011). The high compliance rates observed in the worker sample are broadly consistent with HMRC's operational experience.

- Communication had no effect on student behaviour, and a weak effect on worker behaviour. Making public announcements on the number of individuals caught also had a negligible effect on behaviour for both groups.

- In terms of behavioural responses to the absence of information about fines or audit rates, there was a marked difference between students and workers. The compliance level of students was highest when the audit rate was unknown, while the compliance rate of workers rose when the fine rate was unknown.

- Students and workers shared a number of factors that influenced their decision to under-declare their earnings. On one hand, they were more likely to fully declare their earnings if they were audited in the early stages of the experiment. On the other hand, the more earnings they accumulated in the experiment, the more likely they were to evade.

- Our findings are consistent with the idea that particular social norms which emerge from experiences in the labour market are important drivers of decision-making. We conjecture that a large proportion of our worker sample had internalised a norm of full compliance and that this was carried into the experimental setting. In contrast the student subjects had not internalised this norm and played the experiment as if it were a 'game'.

- Potential explanations for the difference between the two samples include the different recruitment processes for the two different samples; the fact that most of our worker sample took part in the experiment online, as opposed to the lab and different show-up fees and incentives (although these reflect the different opportunity costs of time for each sample). Students may also have been less sensitive than workers to the explicit tax framing used in the instructions.

- Alternatively, we can view our worker sample, who pay their income tax through PAYE, as attitudinally compliant, given that non-compliance is not possible in their daily experience. In contrast, students may approximate the behaviour of taxpayers who are not attitudinally compliant, and as such may provide useful and policy-relevant insights into tax compliance.

1. Introduction and Background

1.1 This report describes an experiment on tax compliance behaviour undertaken online and at the Finance and Economics Experiments Laboratory at Exeter (FEELE) at the University of Exeter between January and April 2012. The research was funded under contract by HM Revenue and Customs.

1.2 The most recent HMRC estimate of the tax gap for Income Tax, National Insurance Contributions, and Capital Gains Tax was £14.4 billion in 2010-11, which is about 45% of the overall tax gap (HMRC, 2012). It is important to understand the processes that motivate and drive the compliance decision in order to design an operational approach that is effective in reducing the tax gap.

1.3 HMRC is a key policy stakeholder in the UK through its role in as a tax collector and administrator of the tax system. Part of the HMRC remit is to assess regularly how effective it is at collecting taxes, which includes studying how different initiatives and interventions can increase tax compliance and therefore reduce the tax gap. This includes understanding the behavioural consequences of particular policies, such as sanctions, which can range from a fine to a criminal prosecution, or communication with taxpayers.

1.4 HMRC aims to increase voluntary compliance through the effective use of sanctions against those who abuse the tax system and evade tax. In particular, HMRC is interested in exploring the deterrent effect that sanctions may have. It is particularly interested in understanding how social networks can aid in the dissemination of information about HMRC auditing strategy and therefore act as catalysts for the deterrent effect. The experiment discussed in this report is designed to investigate the effect of social networks on compliance.

1.5 The initial theoretical work in economics on tax compliance assumed that taxpayers balance the benefits of evading against the penalties they incur if caught. The taxpayer makes an assessment of the probability of being caught and evaluates the payoffs using an expected utility function. This theory, exemplified by the *Yitzhaki model* of the compliance decision, made very clear predictions about the effect on behaviour of key factors, such as the likelihood of being audited, the fine for under-reporting, and the aversion of individuals to taking risks.

1.6 In order to assess the predictive power of this theory public economists have employed laboratory experiments. When using administrative data, one must infer the rates of tax evasion in the economy from the frequency of evaders in the subset of taxpayers that were audited. The reliability of such data decreases as the percentage of taxpayers audited is reduced. In the lab, one can observe directly the actual frequency of tax evasion. One can also systematically vary key factors, such as the audit rate or the tax rate and measure the effect – something which is difficult to achieve in the real world.

1.7 Experiments have become a useful tool for policy-makers in the area of tax compliance. In particular the US Internal Revenue Service has recently commissioned a number of experiments (Alm et al., 2009, Alm et al., 2010) to complement the traditional econometric analysis of tax returns.

1.8 A typical economic experiment is a highly stylised and simplified version of the problem being studied. The methodology ensures a focus on the impact of incentives by simplifying the problem faced by participants and establishing salient financial consequences for the decisions taken by participants. This way, one can

estimate how relevant incentives (e.g. tax rates; penalties for evasion) and institutional features (e.g. auditing rules; social networks) determine behaviour in a very clean way.

1.9 However, the main strength of the methodology can also be seen as its main weakness. By working in a stylised, simulated environment, there are features of the real world which cannot be transposed into the laboratory. For example, one cannot replicate in the laboratory the legal process of prosecuting a tax evader as this would greatly complicate the experiment. When designing an experiment, one always faces a trade-off between a realistic setup and a clean test of hypotheses.

1.10 The traditional expected utility theory of individual compliance behaviour makes a number of very clear predictions. Some of these predictions (such as the effect of an increased fine for non-compliance) are supported by experimental evidence. Other predictions are not, as discussed more fully below.

1.11 The behavioural models which have been developed to explain the evidence incorporate a wider range of factors (e.g. social norms). As a consequence, the predictions of the models do not have the precision of the expected utility theory. Behavioural economics gives a different perspective on the role of experiments. This is that experiments should not be seen as testing the predictions of any particular model but instead have the role of eliciting actual behaviour. The models can then be developed to provide explanations and interpretations of that behaviour.

1.12 Fonseca and Myles (2012) provide a review of the structure and outcome of previous experiments on tax compliance. There are two aspects of past experiments that are important when assessed from a behavioural viewpoint: the choice of subject pool, and the way taxable income is generated.

1.13 First, the vast majority of experiments have been conducted using students as the subject pool. Furthermore, the few experiments with non-student subjects have tended to use teachers or other subjects closely linked to an educational environment (Fonseca and Myles, 2012). One key aspect of behavioural economics is the idea that social rules of behaviour can drive choices. A good experiment design will ensure that these social rules are carried into the experimental setting. This provides a strong argument against using students as subjects: they have no tax-paying experience so may not have internalised the social norm, and will therefore reveal strategic behaviour in the experiment rather than real behaviour.

1.14 Second, previous experiments have generally allocated income to subjects (Fonseca and Myles, 2012). Later experiments have been improved in this respect by introducing tasks through which income is earned. This is important, since the idea of ‘ownership’ of income after expending effort motivates a wish not to lose it easily. Using earned income increases the likelihood that the experiment will reveal actual behaviour rather than strategic behaviour.

1.15 These observations motivate an experimental design in which income is earned by undertaking a task that involves a degree of effort. Our experiment also allowed for subjects to participate online from their choice of location rather than having to enter into the laboratory. Importantly, the subject pool was evenly divided between students and individuals in full-time employment, and therefore taxpayers. These features allow a test of whether the behaviour of students in compliance experiments is a good reflection of the behaviour of non-students. They also permit the effect of the environment (laboratory or outside) to be assessed.

1.16 The remainder of this report is set out as follows:

- Chapter 2 discusses the theoretical background to the experiment with a special emphasis on the implications of behavioural models of the compliance decision.

- Chapter 3 outlines the experimental design.

- Chapter 4 describes the method of sampling and the experimental procedure.

- Chapter 5 provides a statistical analysis of the data collected by the experiment.

- Chapter 6 discusses the results of the experiment and concludes the report.

1.17 A series of appendices are attached to report. They contain the experimental materials (Appendix A), the recruitment materials (Appendix B), auxiliary statistical tables and regression outputs (Appendix C), comments by four independent reviewers (Appendix D), and a literature review (Appendix E).

2. Theoretical Framework

2.1 The economic analysis of the tax compliance decision began with Allingham and Sandmo (1972) who modelled the taxpayer as choosing the level of evasion to maximize expected utility where risk arose from the possibility that an audit might be conducted by the tax authority. Yitzhaki (1974) modified the model to reflect the fact that the punishment for evasion is typically a fine levied on unpaid tax.

2.2 The value of the Yitzhaki model is the clarity of its predictions. The chosen level of evasion will fall when either the penalty rate or the probability of being caught when evading are increased. However, when confronted with values of the audit probability and the fine rate close to those observed in practice, the model predicts that all taxpayers should engage in evasion. This is contradicted by evidence (Fonseca and Myles, 2012). The model also predicts that the taxpayer will declare more income when the tax rate rises. This result has been frequently questioned by empirical evidence, experimental evidence, and intuition.

2.3 There have been numerous extensions to this model to try and explain the difference between predicted and actual levels of compliance. Examples include allowing taxpayers the choice between employment in a formal and informal sectors, and increasing the complexity of declaring income tax (see the surveys of Pyle,1991; Sandmo, 2005) but the basic results are consistent across studies.

2.4 The Yitzhaki model assumes that the preferences of the taxpayer can be represented by an expected utility function and that the taxpayer makes an entirely individualistic decision. In other words, the taxpayer has pre-defined preferences over tax compliance, which correspond to risk attitudes, and does not take into account any social norms, or social context when making her decision. The identification of a range of ‘anomalies’ – observed choices that do not match the predictions of standard theory (see Thaler, 1994) – has inspired the development of behavioural models which encompass many non-standard approaches to choice theory. A growing number of papers have applied the ideas of behavioural economics to the tax compliance decision (see the survey of Hashimzade et al., 2012).

2.5 Two features of these behavioural models are important. First, they can incorporate a decision weight (or ‘belief’) about the likelihood of an audit that can differ from the objective probability. It is not even necessary for the objective probability to be known to the taxpayer. Second, they can include social interaction between taxpayers in the decision problem.

2.6 There are numerous ways in which social interaction can be introduced into the compliance decision. One way is illustrated by the social custom model of Myles and Naylor (1996). Assume that there is a social custom that rewards compliance so that an honest taxpayer receives additional utility which is an increasing function of the proportion of taxpayers who do not evade. This captures the feature that evasion will cause more social prestige to be lost when the non-compliant taxpayer is more out of step with the remainder of society.

2.7 This model makes two important points. The first is that the model can have complete compliance even when the expected financial payoff from evasion is positive. This occurs when the gain from evasion is not sufficient to offset the loss from not following the social custom. The second is that the choice of whether to evade or not depends on the proportion of the population who are evading. The

choice is not made by the taxpayer in isolation but is the outcome of a process of social interaction.

2.8 The early experiments on compliance were focussed on varying the values of factors such as the audit rate to test the predictions of the Yitzhaki model. This model is very precise about the effect of variations in the probability of audit provided that taxpayers know the probability. The behavioural models show that a much broader range of outcomes may occur.

2.9 The behavioural models assume that taxpayers have a subjective belief about the probability of being audited. This makes it acceptable to view the belief of the taxpayer as a characteristic that is carried into the experiment from external experience. If it is, then the effect of treatments that vary the value of the probability of audit will have little consequence on experimental behaviour since there may be resistance to changing a belief. If this is the case, experimental treatments that vary the objective audit probability may not yield significantly different results.

2.10 Similarly, the Yitzhaki model predicts an important effect for communication between taxpayers if it is informative about the audit probability and the fine rate. In contrast, communication may have limited effect on a subjective belief or on a social custom (Crawford, 1998). The observation that many taxpayers choose not to evade can be explained by the existence of a social custom of honest payment. If the social custom is brought into the experiment from outside and evolves slowly over time, it may not adjust over the time horizon of an experiment. This can provide another reason why compliance behaviour may not respond to treatments.

2.11 The Yitzhaki model with expected utility predicts that there should be no difference in the behaviour of students and non-students in a compliance experiment. Behavioural models are consistent with a very different outcome. Those with experience of tax payment will have had an opportunity and reason to form beliefs about the probability of audit, and the punishment if caught. They will also have been involved in the socialization process through which taxpayers absorb the social custom. This opens the possibility that the two subject groups could have very different behaviour in the experimental setting.

3. Experimental Design

3.1 We utilise an experiment with a 5-stage design throughout the project. The experimental design is partly new, but it is sufficiently similar at a conceptual level to the existing body of experimental research on tax evasion to permit comparison of the findings to the existing literature.

3.2 We outline the basic experimental procedure first, and then proceed to the different experimental treatments that result from the different ways that each task can be executed. The experiment focuses on a key area of interest for HMRC, the effect that social networks have on taxpayer compliance[1].

3.3 However, the experiment also examines how other factors (awareness of auditing rates, fines, nature of communication, and severity of fines) affect compliance behaviour by using different treatment conditions. We conclude the section by reporting how we executed the experiment for both student and worker samples.

3.4 The experimental design involved three main parts. The first part consisted of the main experiment testing tax compliance. The second part was a risk aversion elicitation experiment. The third part consisted of a number of questionnaires eliciting personality measures. We describe each one in turn.

3.5 In all parts of the experiment, participants made decisions that had financial consequences. As the experiment progressed participants accumulated a payoff that was expressed in Experimental Currency Units (ECUs). The ECUs were converted to financial payments at the end of the experiment using an exchange rate known to participants in advance.

Part 1: Tax Compliance Experiment

3.6 The five stages of the experiment were as follows:

1. Income generation
2. Declaration of income for tax purposes
3. Auditing
4. Payoffs for round
5. Communication among taxpayers via a social network.

3.7 A period in the experiment consisted of Stages 1-5 above. In every experimental session participants repeated the play for 15 periods. The reason for multiple periods within the experiment is that as participants become more familiar with the experimental setup, they can be expected to adapt and fine-tune their behaviour. This means that there needs to be multiple iterations or periods in the experiment to allow for a process of learning.

3.8 The rationale for choosing 15 periods is to maximise the amount of data we obtained for each participant, while keeping the duration of the experiment within reasonable limits. It is standard practice in experimental economics to ensure that experimental sessions last for no longer than 2 hours. The duration of an experiment is an important consideration for two reasons. Firstly, it directly determines the level of incentive that must be paid to subjects and therefore the cost of running a session.

[1] See section 1 for more details.

Secondly, a lengthy experiment is likely to lead to subject fatigue, which in turn leads to a deterioration of data quality.

3.9 It was initially intended that a 'Stage 0' in which subjects were allocated to types, such as employment status or demographic characteristics, would take place before the start of the experiment. We chose not to implement this because it was not possible to ensure sufficient variability in each group/session to make the type distinction meaningful. We opted not to classify respondents into artificial types as the usefulness of this was not apparent. As an alternative we chose instead to exploit individual-level heterogeneity through ability via a 'real effort' experimental task, which is described below.

3.10 The individual stages of the game were:

Stage 1: Real-effort task
Participants performed a real-effort task for a fixed piece rate. Performance in this task generated the income of a participant in each period. The task involved expenditure of time and effort and the outcome was dependent upon the level of skill. It was designed to induce a feeling of "ownership" of income. The task consisted of a set of 48 sliders on a screen (Gill and Prowse, 2012). Participants earned a fixed payoff by placing a sliders at its halfway point. For each slider solved, participants would earn a payoff of 1 ECU.[2] The total payoff – the income level for that round – was determined by the number of sliders correctly placed at the halfway point. A screenshot of the task is in the sample instruction sets in Appendix A.

Stage 2: Tax declaration
Participants were told their income from Stage 1 and had to declare their taxable income. This involved inserting a number into a simplified tax return.

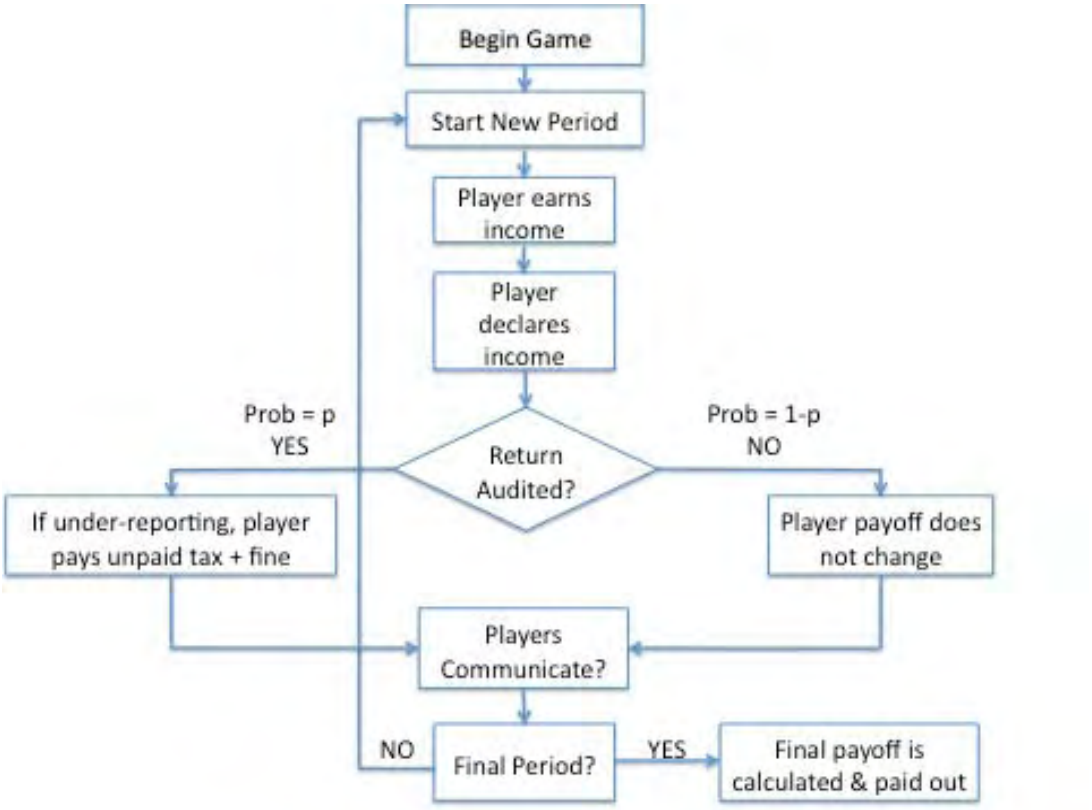
Stage 3: Auditing
Participants were audited with a fixed probability *p*. If an audit occurred and subjects truthfully declared their taxable income, no penalty was levied. Otherwise, subjects paid a fine *f*.

Stage 4: Final Period Payoffs
Participants received final payoffs for the round.

Stage 5: Communication
Information about fines and auditing frequencies was transmitted to/by participants. How, and by whom, this information was transmitted was a treatment variable.

[2] Each ECU had a pre-defined exchange rate with sterling. See section 4 for more details.

Figure 1: Experimental Procedure



3.11 We now outline the different treatments that form the basis of the experiment. These treatments seek to address each of the research questions proposed by the HMRC.

3.12 **Research Question 1: How do social networks affect compliance?**

Research Question 2: How does publicising information about numbers of evaders who have been caught affect compliance?

We interpreted a social network as a collection of individuals in which social interaction is easier, such as a trade association, and in which communication about tax auditing probabilities and fines is likely to occur. We took groups where no communication between the players was allowed (NOCOMM) as the baseline condition. This is the condition against which we measured the effectiveness of our treatment conditions. We considered two conditions in addition to NOCOMM, which were different variants of Stage 5 where the players communicate:

- NOCOMM: no communication was allowed.
- FREE: communication was unrestricted between players. Each subject could communicate with anyone in the 5-strong group using a chat box similar to Skype.
- RESTRICTED: subjects could only communicate with 2 other players using the chat box.

3.13 We addressed Research Question 1 by comparing compliance behaviour in NOCOMM to compliance behaviour in FREE and RESTRICTED, where participants were allowed to communicate with each other. In the former, everyone in a 5-person group

could communicate freely, which allowed for a quick dissemination of information. In the latter treatment, since each player could only communicate with two others, the flow of information should be slower, and the effectiveness of communication should be reduced.[3]

3.14 The communication treatments approximated the type of communication taxpayers may have with people within their social or professional network. We opted for free-form communication as it is a more natural environment in which participants can discuss what they feel is important to inform their decisions, rather than imposing a structured form of communication.

3.15 We addressed Research Question 2 by comparing compliance behaviour in NOCOMM to compliance behaviour in another variant of Stage 5, in which the experimental authority made public announcements about compliance rates and fines issued to non-compliant participants.

- PUBLIC: the computer (who played the role of the tax auditing agency) announced to all subjects the number of audits and total fines collected at the end of each period. No communication was allowed between the players.

3.16 **Research Question 3: How does awareness of a tax authority's audit rate affect compliance?**

We considered the following conditions, which manipulated information about the likelihood of auditing and type of auditing rule in Stage 3.

- Known Probability (KP): Each taxpayer was audited with probability *p* in every period, which was public knowledge.
- Unknown Probability (UP): Each taxpayer was audited with probability *p* in every period, which was not known to taxpayers
- Unknown Probability, Follow-Up (UP-FU): Each taxpayer was audited with probability *p*, in every period, which was unknown to taxpayers. If caught, *p* = 1 (i.e. the player was certain to be audited) for the next 2 periods.[4]

3.17 The Unknown Probability condition tests for the effect of unknown audit rates, compared with the condition where audit rates are known. The former approximates reality in the sense that most taxpayers are unlikely to know what HMRC's true audit rate is. The Follow-Up condition tests for the effectiveness of the threat of a follow-up audit for non-compliant participants. We implemented this follow-up rule only when the audit rate was unknown, as it is the more realistic scenario.

3.18 **Research Question 4: How does awareness of audit outcomes (fines) affect compliance?**

We considered two conditions, which manipulated information about fine levels.

[3] We discovered upon programming the experiment that there were serious limitations in the ability of the software used to cope with the real effort task where groups had more than 5 players. This forced us opt for a simplified experiment design where a group of five people was an independent observation. Social network effects were defined in terms of whether subjects could communicate with the remaining four participants in the group (FREE) or whether they could communicate with only a subset of the participants (RESTRICTED). The latter case simulated a network shaped as a circle, where each node of the network could only communicate with its two adjacent nodes.

[4] In all cases, the audit rule was common knowledge among participants.

- Known Fine (KF): An under-reporting taxpayer who is audited pays fine *f*, the rate of which is public knowledge.
- Unknown Fine (UF): An under-reporting taxpayer pays an unknown fine *f* if audited.

3.19 This led to the following experimental design (ticks denote treatments that were executed, crosses denote treatments which were not relevant.) We term this group as *Treatment Set* 1.

Table 1: Design of Treatment Set 1

	KNOWN P KNOWN F	KNOWN P UNKNOWN F	UNKNOWN P KNOWN F	UNKNOWN P, FOLLOW UP UNKNOWN F
FREE	×	√	√	√
RESTRICTED	×	√	√	√
PUBLIC	×	√	√	√
NOCOMM	√	√	√	√

3.20 **Research Question 5: How do the nature and severity of penalties affect compliance?**

To answer the fifth research question we introduced new treatments to test the impact of changing the size of the fine levied on under-reporting subjects (HIGH F (200% of the tax evaded) and LOW F (100% of the tax evaded)) and the effect of changing the auditing and two audit probabilities (HIGH P (40%) and LOW P (20%)).

3.21 These treatments had two purposes. On one hand, it tackled the important question of determining declaration-auditing and declaration-fine elasticities. On the other hand, it was useful as a calibration exercise for the first experiment. This was conducted solely in the no communication setting and complete information – the NOCOMM KP KF condition.[5] We term these *Treatment Set* 2.

Table 2: Treatment Set 2 Design

	LOW F	HIGH F
LOW P	√	√
HIGH P	√	√

3.22 The choice of the specific audit probabilities was to ensure that there was a sufficiently high probability that each taxpayer would be audited at least once during the experiment. With only 15 periods, we felt that lower audit rates would not form a credible deterrent. The choice of fine levels was in line with the current policy operated by HMRC.

3.23 The experimental design is a stylised model of the real world. It is clear that it provides a simplified representation of how taxpayers make their tax declarations and how HMRC monitors and pursues non-compliant taxpayers. This is to facilitate participants' understanding of the task thus increasing the quality of the data, and to focus on the relevant research questions. As such, there are obvious limits on the extent to which we can generalise the results from the lab to the real world.

[5] Also note that the NOCOMM KPKF condition in Treatment Set 1 is the same condition as LOW P LOW F in Treatment Set 2.

Part 2: Risk Aversion Elicitation Experiment

3.24 The second part of each session used a test designed to elicit individuals' risk attitudes. We employed the design of Holt and Laury (2002). This experiment measures how averse individuals are to financial risk by presenting participants with a series of scenarios.

3.25 Each scenario involved two gambles, and each gamble had two possible payoffs. One gamble had one very high payoff and a very low payoff (e.g. 77 and 2). The other gamble had two intermediate-sized payoffs (e.g. 40 and 32). By systematically varying the probability of obtaining the high payoff, we were able to determine how averse to risk each participant was. See Appendix A for details of this procedure.

Part 3: Personality Survey

3.26 The final part of each session consisted of a set of questionnaires aimed at obtaining measures of personality. There is a consensus among researchers in psychology that a comprehensive description of an individual's personality rests upon five broadly defined factors: Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Openness (John and Srivastava, 1999; McCrae and Costa, 1999).

3.27 The role of each of these factors is as follows:

- Extraversion looks at one's preference for social contact;
- Agreeableness embodies the tendency for compassion and cooperation;
- Conscientiousness reflects the ability for self-discipline, organization and ability for planning behaviour;
- Emotional Stability captures the degree of stability and impulsivity in emotional responses;
- Openness captures the degree of intellectual curiosity, creativity, and a taste for novelty;

3.28 This 'Big Five' model is an empirically derived construct of how human personality can be categorised in social psychology. Alaheto (2003) found that individuals who had high scores on the extraversion and emotional instability, as well as low scores on agreeableness, were more likely to commit white collar crimes, of which tax evasion was an example. As such we included a questionnaire on the big five, to understand to what extent Alaheto's findings extend to our data set.

3.29 The measurement of these five factors requires extremely extensive questionnaires which can have up to 240 questions (Costa and McCrae, 1992). A questionnaire of this length was inappropriate for our purposes. We opted for a questionnaire developed by Woods and Hampson (2005) which only encompasses one question per personality factor. This questionnaire provides two descriptions for each personality factor and asks individuals to rate how close is each description to what they believe they are like. Details of the questionnaire are provided in Appendix A.

4. Experimental Procedures

4.1 We will describe the experimental protocol used for students and workers separately, as the setup for the two subject pools differed in important ways. The common element in both protocols was the fact that participants faced the same computer interface.[6]

4.2 Participants read the instruction set on the computer screen for 10 minutes before the experiment started.[7]

4.3 The fact that the software was the same in both protocols meant the duration of the experiment was roughly the same, conditional on the whether communication between players was allowed - FREE and RESTRICTED treatments meant that these experiments were necessarily longer than PUBLIC and NOCOMM.

4.4 Payoffs in the experiment were denominated in a fictitious currency, the Experimental Currency Unit (ECU). This was done in order to keep the same numerical parameters in the experiment for both participant pools. Since workers require stronger monetary incentives than students because of their higher opportunity cost of time, we implemented two different ECU/pound exchange rates. In the student sessions, the ECU/pound exchange rate was 30/1; in the worker sessions the exchange rate was 15/1. Sessions lasted on average 75 minutes and never longer than 90 minutes.

4.5 The experiment involved a total of 1,023 participants. 500 of the participants were students and 523 were workers. Details about the breakdown of participants per condition are given in Tables 1 and 2.

1. Student Sample

4.6 We recruited the student sample from a database of registered participants.[8] The participants were undergraduate students at the University of Exeter. The database sent an email to a random set of participants, who had not taken part in any similar experiments. Participants signed up to their preferred time slot, and received an email reminder 24 hours before their session took place. To avoid having too few participants in a given session, we overbooked sessions -- for a 20-participant session, we typically sent out 24 invitations. Participants who arrived on time but could not participate received a show-up fee of £5.

4.7 The experiment took place at the FEELE laboratory at the University of Exeter Business School. The laboratory is a dedicated room with 20 networked computers in booths separated by physical partitions. These partitions do not allow for participants to see into others' booths, and severely constrain communication between participants. All interaction between participants is done through the software provided. That, in addition to the fact that several groups of five played at the same time, reinforced the anonymity of participants.

4.8 Upon arrival at the laboratory, participants were assigned to a booth. When all participants were seated, the software was started and the instructions were shown on the computer screen. To minimise the procedural discrepancies with the

[6] The experiment was programmed using z-Tree software (Fischbacher, 2007).
[7] A sample copy of one of the instruction sets in available in Appendix A.
[8] We used the database management software ORSEE (Greiner, 2004).

worker participant pool sessions, the experimenters did not take any questions from participants during the session.

4.9 Upon completing the experiment, participants were paid the sum total of their earnings in parts 1 and 2 in cash. A total of 25 sessions took place in January and February 2012 and a total of 500 student participants took part in the experiment. The average payment was £15.89, which included a show-up fee of £5.

2. Worker Sample

4.10 The worker sample was divided into participants who did the experiment online, and participants who travelled to the laboratory. The rationale for using online experiments was twofold: firstly, it could provide us with a nationwide subject pool, which was more representative than an Exeter-based sample, or in fact for any one location in the UK. Secondly, it also allowed our subjects to engage in the experiment in a more natural environment, such as their homes or place of employment. The online environment also afforded participants with a greater degree of anonymity relative to each other. We were able to test whether this was an important factor in determining behaviour (see section 5).

4.11 We recruited part of our worker sample through a survey company, Saros Research. Saros' database consisted of a sample of adult UK residents, who had previously registered their interest in taking part in market research activities. We cannot guarantee that this sample is fully representative of the UK's population with regards to income distribution and/or occupational activity, since our participants were drawn from a population of volunteers.[9]

4.12 We screened participants for eligibility in our study on the basis of two main criteria. Firstly, participants had to be full-time UK residents for tax purposes. Secondly, participants had to be able to download and install the software used to run the experiment.

4.13 The second criterion introduces a sample selection bias: participants had to have an Internet connection, own a Windows-powered computer, and be able to install and run our software. While all the individuals in Saros' database met the first criterion, some did not own a Windows PC. We tried to minimise the third potential for sample bias by creating a simple 3-step process to download and install our software (a copy of the instructions to participants is in Appendix A.) Nevertheless, this may have discouraged some people from participating. We however do not believe this sample selection bias affected our results, since we have no reason to believe that computer proficiency (or use of Windows operating systems) is correlated with tax compliance behaviour.

4.14 The screening process consisted of an extensive survey, which collected information on the two aforementioned criteria; the survey also collected voluntary information on age, gender, annual personal income, annual household income, industry, work experience, marital status, area of residence, education level, and ethnicity. However, we did not obtain a complete set of observations for all participants, which limits our ability to make inference on all the socio-economic variables of interest.

[9] Note that the same criticism applies to student samples, which may not be representative of the university population. See Cleave et al. (2012) for a study of sample selection bias in lab experiments.

4.15 Upon collecting a set of eligible participants, we invited participants via email to take part in a session. Most sessions were scheduled between 18:00 and 22:00 during working days and throughout the day on Saturdays and Sundays, in order to ensure take-up (although a few sessions were run between 12:00 and 17:00 on some working days.)

4.16 Upon registering for a session, participants received a confirmation email, which included a link that enabled participants to download a bespoke file, which when run at the designated date and time, linked their PC to the server at the University of Exeter. Participants were also asked to register their details on a secure server, which consisted of an address should they wish to get paid via cheque, or their bank details should they prefer a bank transfer.

4.17 To ensure that participants managed to download the software correctly, as well as to ensure a good show-up rate, we sent reminder emails on the day of the session. We also made telephone calls to participants approximately an hour before the commencement of a session to ensure they had received our reminders and had had no difficulty installing the software.

4.18 Participants were asked to log on a few minutes before the session's scheduled start time. The server software was initiated one or two minutes after the scheduled start time to allow for any connection problems and latecomers. Upon completing the experiment, a message appeared on the screen asking participants to close the program, and informing them their payment would be processed in the days following the session. Participants were paid typically within 5 working days of their session.

4.19 Despite these efforts, we had a show-up rate of about 80%. This included participants who cancelled, as well as those who decided not to continue the session upon reading the instruction sets. Since each bespoke file connected a participant to a given server, which in turn managed a group of five participants, this meant we could not overbook sessions. As such, we treated all inactive participants (i.e. no-shows and drop-outs) as instances of participants who did not exert any effort and therefore generated no income. We controlled for the effect of group size in the econometric analysis.

4.20 The high no-show rates in our online sessions forced us to recruit non-student participants locally. We also recruited non-students through email to local Exeter residents as well as employees of the University and local businesses. Participants were invited to take part in a decision-making experiment, which would last up to 90 minutes and which guaranteed a show-up fee of £20. They were invited to register their interest in participating online, and they received a confirmation email with the location of the lab, as well as the date and time of session. Participants were paid in cash at the end of the session. They also filled in an additional questionnaire at the end of the session containing the same socio-demographic questions as those asked by Saros when screening for their participants.

4.21 Sessions took place in March 2012 and a total of 523 participants took part in the experiment, 118 of which took part in Exeter lab sessions. The average payment was £36.03, which included a show-up fee of £20. Once the data collection was complete, we emailed all participants a debrief document (in Appendix A), which outlined the objectives of the research, as well as the project's source of funding. Participants could opt to have their data deleted from the project by filling in a short form and posting it to the FEELE Lab. We received no such requests a month after completing the data collection.

4.22 We suffered a number of computer-related issues that hindered our data collection. We were not able to collect final period data, risk aversion and questionnaire information for 21 groups (9% of the sample). This led to us dropping observations whenever we included information on risk aversion as well as socio-economics. We conducted robustness checks whenever possible and we report them whenever appropriate in the analysis section. We did not find any evidence to suggest our analysis is compromised by dropping these observations.

5. Results

Introduction

5.1 The data analysis section is divided in two complementary sections: the first deals with average treatment effects on levels of compliance across groups; the second models participants' compliance decision using more sophisticated econometric techniques. We begin by making a number of observations regarding the data analysis.

5.2 First, the main unit of analysis will be the rate of compliance. We term this variable throughout the data analysis as the *compliance rate* and define it as the ratio of an individuals' declared income to income earned in a given period:

$$Compliance\ rate = \frac{Income\ declared}{Income\ earned}.$$

This definition implies that the compliance rate has a value between zero and one. This imposes constraints on our method of data analysis when we go beyond analysing average differences between treatments. We elaborate on this issue in the appropriate sub-section.

5.3 Second, we note that we dropped from the analysis observations recording over-compliance where respondents reported more income than they had earned in the real effort task. We assume these instances to be errors by subjects and as such we treat them as outliers. These observations account for about 3% of our sample, so we believe the integrity of the analysis is not compromised.

5.4 The following box provides a summary of the findings regarding each of the research questions underpinning the present report.

Summary of Results

We find three significant differences between our two main samples.

1. Workers have significantly higher compliance rates than students.

2. Students are more responsive to incentives than workers.

3. Workers' compliance is higher when they lack information about fine levels, while students' compliance is higher when they lack information about audit rates.

Research Question 1: How do social networks affect compliance?
Social Networks have very little impact on compliance. Compliance behaviour by both sets of participants (workers and students) is not systematically affected by the ability to communicate with their peers.
We find weak evidence that communication leads to higher rates of tax compliance among workers.

Research Question 2: How does publicising information about evaders who have been caught affect compliance?

Publicising information about evaders has no significant effect on compliance by workers or students.

Research Question 3: How does awareness of HMRC's audit rate affect compliance?

Awareness of the likelihood of being audited leads non-complying students to declare a lower proportion of their income, but it has no effect on the rate of full tax compliance itself. It has no effect on compliance behaviour of workers.

Research Question 4: How does awareness of audit outcomes (fines) affect compliance?

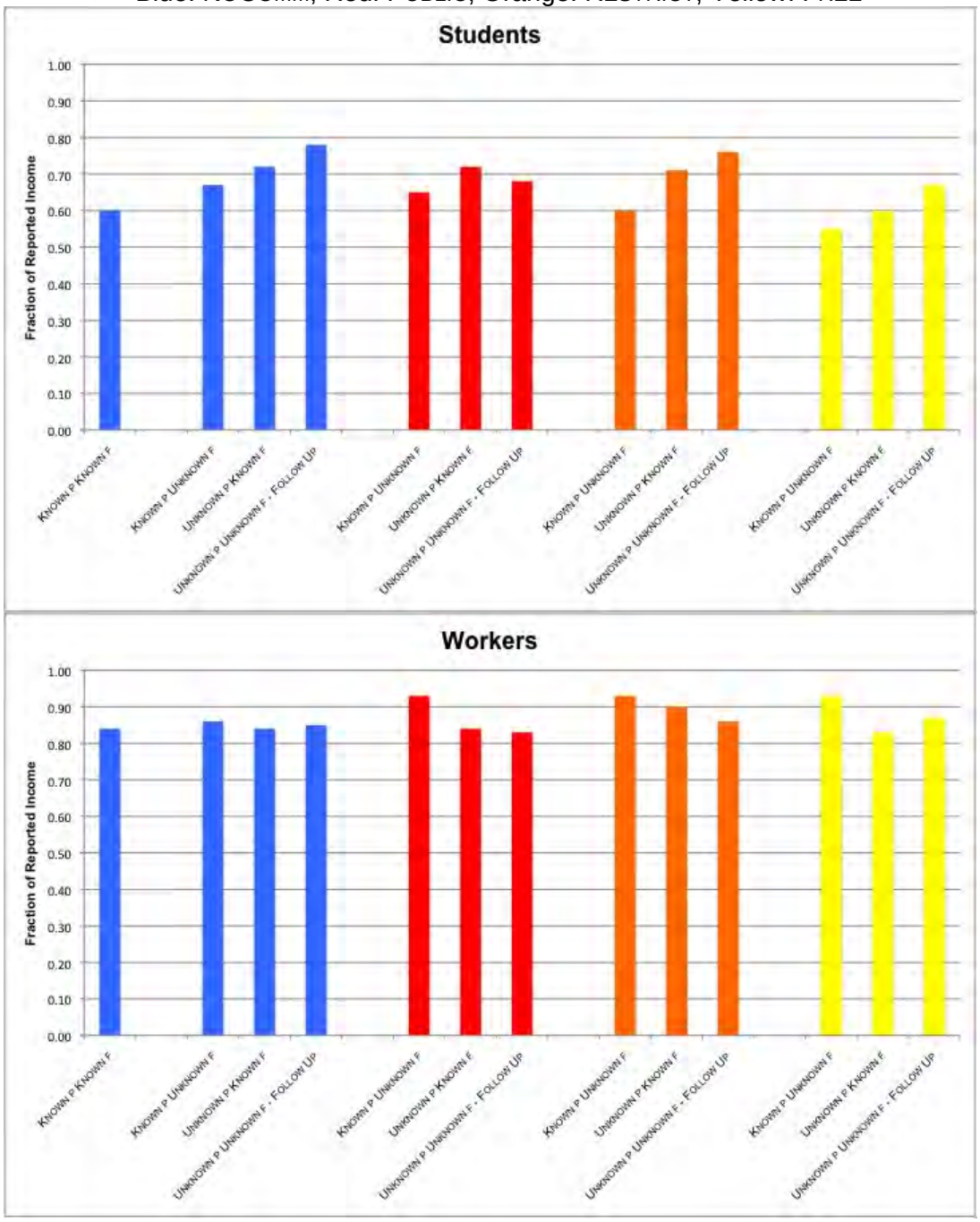
Awareness of the level of fines levied on non-complying taxpayers leads to a decrease in the rate of full compliers in workers, though it has no effect on the proportion of income declared by those who do not comply.
It has no effect on the compliance behaviour of students.

Research Question 5: How do the nature and severity of penalties affect compliance?

Changing the severity of fines in isolation only had a significant impact on student behaviour, in that it led to a 12-percentage point increase in the likelihood of full compliance.
It had no impact on worker behaviour.
Increasing the audit rate in isolation had no impact on either student or worker compliance behaviour.
Increasing both audit rates and fines simultaneously led to a 20-percentage point increase in full compliance by students and an 11-percentage point increase in workers.

Summary Statistics

Figure 2: Fraction of Reported Income: Treatment Set 1. Colours indicate treatment. Blue: NoComm, Red: Public; Orange: Restrict; Yellow: Free



5.5 We begin the data analysis by looking at simple average treatment effects between *groups*. We apply standard statistical tests using the average compliance rate at the group level over 15 rounds as the unit of analysis.[10] This is because

[10] We use the Mann-Whitney test for equality of medians. This test is recommended whenever the data under scrutiny may not follow a normal distribution and/or variance of the distribution is not homogeneous. Either case renders the t-test invalid. We also report significance levels of 5% or lower, which mean we can be can be 95% confident that the differences observed in the sample can be generalised to the population, assuming a similar environmental setting.

observations in later rounds are not independent of earlier rounds – either because of learning, or due to strategic interactions between individuals through the chat conditions.[11]

5.6 Figure 2 displays the average fraction of reported income in Treatment Set 1 for all information conditions and both participant pools. It is immediately noticeable that in every case the compliance of workers is greater than that of students.

5.7 We now look at the impact of information about the rate of fine and/or the probability of audit as well as how the ability to communicate (or receive information) about these variables affected compliance behaviour.

5.8 In the baseline case where the audit and fine rate were known, and no communication was allowed (NOCOMM KPKF), the average compliance rate in the worker sample was 0.84 (84%). Keeping communication constant, we find that withholding information about fines has no statistically significant impact on average compliance. For the student sample, the baseline rate of compliance was 0.60 (60%). There was evidence of a statistically significant increase in the average compliance rate only when follow-up audits were carried out with certainty when players were caught evading (NOCOMM UP-FU).[12]

5.9 We now focus on the effect of communication conditional when holding the amount of information available to players about the audit and fine rate is held constant. We begin by comparing behaviour in the NOCOMM treatment to each of the communication treatments. There are no significant differences in average behaviour when comparing groups which could and could not communicate with two exceptions. There is a statistically significant difference between:

- No communication and free communication, when the audit rate was unknown and the fine rate was known (NOCOMM UPKF versus FREE UPKF)
- No communication and free communication when a follow-up audit was used (NOCOMM UP-FU versus FREE UP-FU).[13]

5.10 Figure 3 displays average compliance rates in Treatment Set 2 for both participant pools. The KNOWN P KNOWN F condition represents the case where both the probability of audit and the fine level are at the low level. It is again seen that the effect of changing the audit probability and/or the fine rate differs between the subject pools.

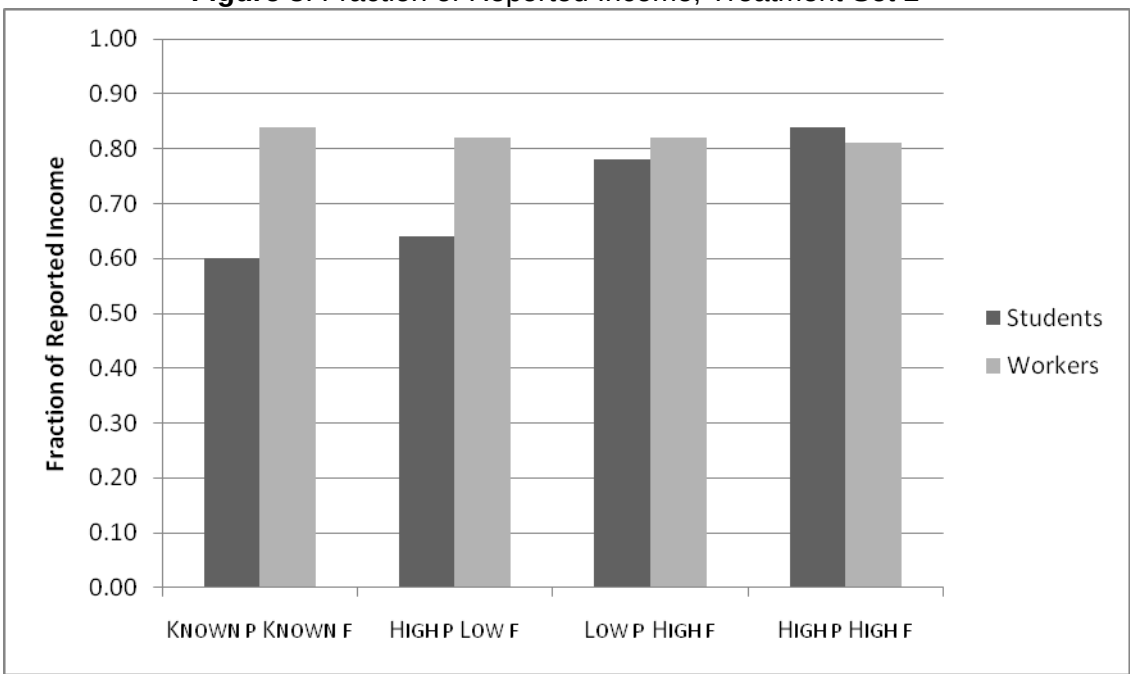
5.11 There is no statistically significant impact of changing either of the variables on the worker compliance rate.[14]

5.12 The average compliance rate in all treatments is always around 80%. Students respond more to changes in the fine rate, rather than to increases in the probability of audit. An increase in the fine rate in isolation leads to average compliance increasing from 0.60 to 0.78.[15] An increase in the probability of auditing leads to a rise in the compliance rate from 0.60 to 0.64 but that rise is not

[11] We complement that analysis with standard Ordinary Least Squares OLS regressions using individual compliance rate in a given round as the unit of analysis. This complementary approach takes advantage of the larger sample size. Results of these regressions are presented in Appendix C (Table 4).
[12] Mann-Whitney U test (henceforth, MWU), $p = 0.036$.
[13] MWU test, $p = 0.047$ in both tests.
[14] Comparing average compliance rate in LOW P LOW F to: HIGH P LOW F: $p = 0.498$; LOW P HIGH F: $p = 0.325$; HIGH P HIGH F: $p = 0.248$, all MWU tests.
[15] $p = 0.020$, MWU test

significant.[16] Increasing both the probability of audit and fines leads to an increase in average compliance to 0.84.[17]

Figure 3: Fraction of Reported Income, Treatment Set 2



5.13 The difference in compliance rates in the baseline condition between students and workers disappears when both the probability of audit and the fine level are both high.[18]

The Compliance Decision

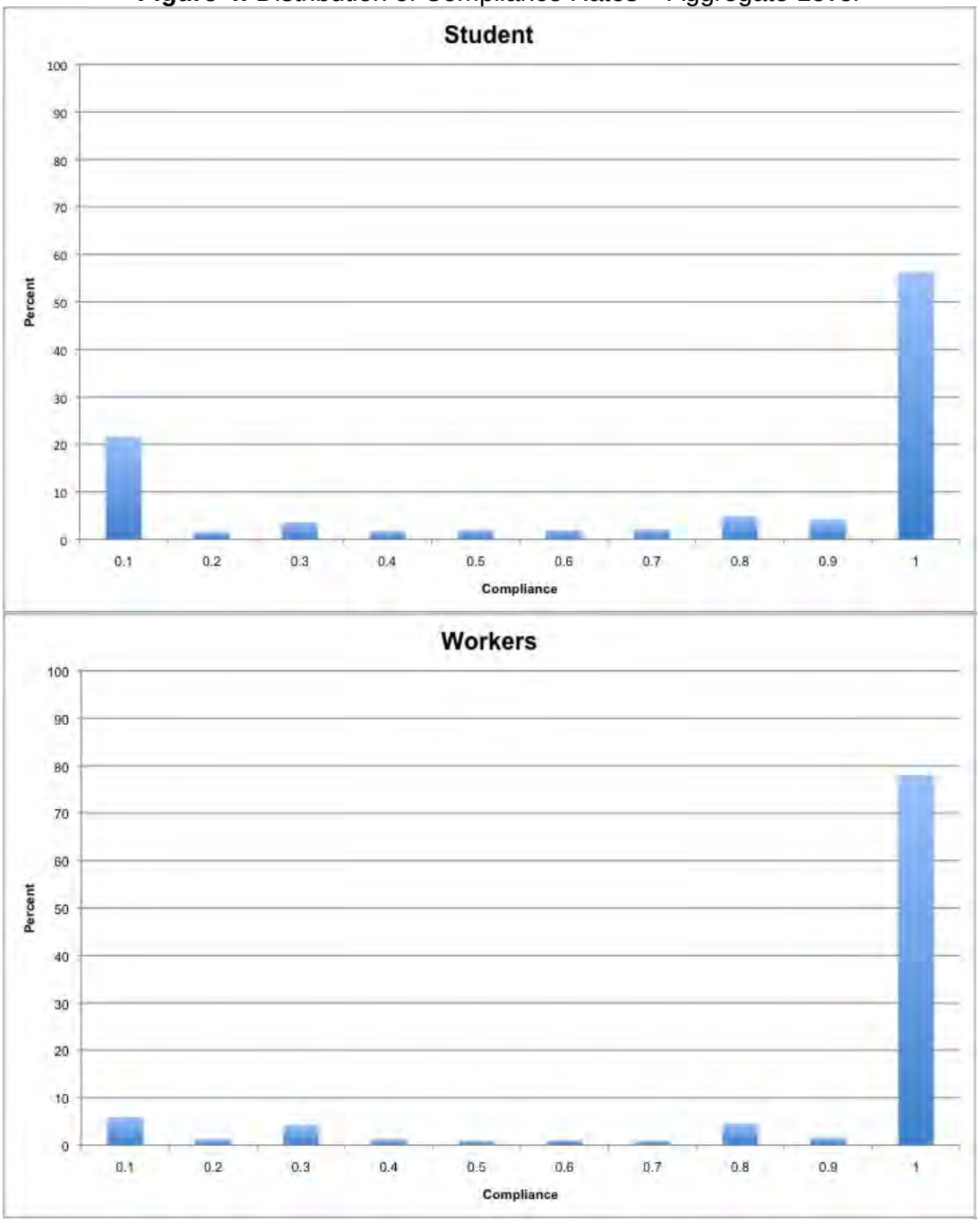
5.14 We now describe an econometric analysis of the compliance decision by participants (students and workers). The aim is to determine the importance of the factors that affect the decision. The compliance variable is a ratio which has a natural lower bound of zero (complete evasion), and a theoretical upper bound of one. As noted, there are outliers where the variable is above one but we exclude the possibility that an individual may deliberately want to pay tax on non-existent earnings.

5.15 There are a non-trivial proportion of observations at the upper boundary point of full compliance, demonstrating that in the majority of cases, participants declared 100% of their income.

5.16 The nature of the distribution of compliance behaviour meant that the most appropriate way for analysing the data was to estimate a two-part model (see Appendix C for more details). Observations at the upper boundary of the support of the distribution (i.e. full compliance) were modelled as a standard binary model, and the remaining observations (i.e. where compliance rates were less than 100%) were modelled using a fractional regression model.

[16] p = 0.650, MWU test
[17] p = 0.001, MWU test
[18] p = 0.393, MWU test

Figure 4: Distribution of Compliance Rates – Aggregate Level



5.17 In short, we formally modelled the features of the distribution of compliance as two distinct decisions. The first decision is whether to evade or not. We treat this as a binary decision (i.e. yes or no). The second part of the model determines how much is evaded conditional on when participants had decided to not declare at least some of their income.

Table 3: Descriptive Statistics

Variable	Observations	Mean (Std. Dev.)	Range	Description
NoComm	15,226	0.45 (0.50)	{0,1}	1 if no communication was allowed
Free	15,226	0.18 (0.38)	{0,1}	1 if participants could communicate with all
Restrict	15,226	0.18 (0.39)	{0,1}	1 if participants could communicate with 2 others
Public	15,226	0.18 (0.39)	{0,1}	1 if public announcement on audits and fines
Known Prob	15,226	0.52 (0.50)	{0,1}	1 if participants knew audit probability
Known Fine	15,226	0.51 (0.50)	{0,1}	1 if participants knew fine level
Ability	15,226	20.06 (9.31)	[0,48]	No. of solved puzzles solved in period
Income	15,226	1831.89 (965.13)	[500,6493]	Total cumulative income (including show-up fee)
Audited & Not Caught (t-1)	14,203	0.25 (0.44)	{0,1}	1 if audited in previous period
Audited & Caught (t-1)	14,203	0.09 (0.28)	{0,1}	1 if audited and caught in previous period
Early Audited & Not Caught	15,226	0.08 (0.28)	{0,1}	1 if audited in first five periods
Early Audited & Caught	15,226	0.03 (0.16)	{0,1}	1 if audited and caught in first five periods
Worker	15,226	0.51 (0.50)	{0,1}	1 if not in education
Lab	15,226	0.60 (0.49)	{0,1}	1 if observation collected in the lab
Group Size	15,226	4.59 (0.72)	[1,5]	Number of participants in a group
Extraversion	13,614	4.54 (2.04)	[1,9]	Categorical classification of extraversion
Agreeableness	13,599	5.44 (2.12)	[1,9]	Categorical classification of agreeableness
Emotional Stability	13,599	4.31 (2.17)	[1,9]	Categorical classification of emotional stability
Conscientiousness	13,599	4.27 (2.22)	[1,9]	Categorical classification of conscientiousness
Openness	13,614	5.47 (2.13)	[1,9]	Categorical classification of openness
Risk aversion	12,241	6.04 (2.39)	[1,9]	Categorical classification of degree of risk aversion
Personal Income	7,885	2.58 (0.70)	[1,9]	Categorical classification of annual personal income bands
Household Income	7,810	2.62 (1.23)	[1,9]	Categorical classification of annual personal income bands
Age	13,946	29.00 (11.82)	[18,68]	Age
Male	13,946	0.50 (0.50)	{0,1}	1 if male
Work Experience	7,630	8.71 (7.62)	[0,37]	No. of years in current occupation

5.18 Our estimation strategy for each subject pool is divided into three nested regressions. The first regression contains only treatment fixed effects, consisting of variables which show the association between the treatments described in section 4 and the compliance decision. Our second regression contains payoff-relevant variables, such as accumulated income up to that round and whether a participant was audited or not. The third regression further extends the analysis to include behaviourally-relevant characteristics at the individual level.

5.19 Table 3 provides a list of all the factors included in the models, and some summary statistics. Model results are presented in Appendix C (Tables 2 and 3).

5.20 Research Question 1 concerns the ability to communicate with members of one's social network. The econometric analysis shows that communications typically increases compliance. However, the manner in which this effect is manifested differs depending on the participant pool.

5.21 Table 4 outlines the estimated change in compliance behaviour associated with each treatment compared to the baseline case, where no communication was possible (NOCOMM). In the Evade regressions, the coefficients represent the change in the probability of choosing to evade vis-à-vis NOCOMM. In the Percentage of Compliance regressions, the coefficients should be interpreted as the change in the proportion of income declared associated with each treatment compared to the baseline.

5.22 In the case of workers, changes in behaviour are manifested mainly through a lower likelihood of evasion – participants who communicate with all group members (FREE) are 6 percentage points less likely to evade than in NOCOMM, though this difference is not significant, and participants in RESTRICT are 8 percentage points less likely to evade than in NOCOMM[19]. However, conditional on evading, communication does not have any effect on the proportion of income that participants report.

5.23 In the case of students, both the likelihood of evasion and the proportion of income evaded are not associated with whether participants could communicate with each other or not.

5.24 In other words, communication seems to affect the decision of workers on whether or not to evade, but only when communication is restricted to two other individuals. This effect works by decreasing the likelihood of evading, suggesting that communication works as a norm reinforcement device. Communication has no such effect on student behaviour.

5.25 Research Question 2 concerned the impact of public announcements from the tax authority concerning non-compliant taxpayers. We find that such announcements have no impact on behaviour, either in terms of the likelihood of evading, or regarding the extent to which people evade (see Table 4).

5.26 Research Questions 3 and 4 concerned the impact of the lack of information about fines or audit rates on compliance behaviour. Table 5 reports marginal effects from the econometric analysis of the probability of evasion and the proportion of income reported, given that a decision has been made to evade. The baseline of comparison is the full information condition, where both the audit rate and fine level are known.

[19] $p = 0.025$.

Table 4: Marginal Effects of Communication Condition on Compliance

Model	Evade	Percentage of Compliance	Evade	Percentage of Compliance
	Student		Worker	
FREE	0.05	-0.07	-0.06	-0.00
	(0.06)	(0.04)	(0.04)	(0.06)
RESTRICT	-0.01	-0.02	-0.08**	0.02
	(0.05)	(0.05)	(0.04)	(0.06)
PUBLIC	-0.01	-0.01	-0.05	-0.03
	(0.05)	(0.04)	(0.05)	(0.06)
Baseline	0.48	0.31	0.22	0.38
N	5,609	2,712	5,949	1,258

Group-level clustered standard errors in parenthesis.
***, **: significant at the 1% and 5% level

5.27 For the worker sample, Table 5 shows that not disclosing fine information while keeping audit rates known reduces the likelihood of evasion by 16 percentage points.[20] However, for those participants who choose not to declare their full earnings, there was no difference in compliance relative to the group with full information.

5.28 Not disclosing information about audit rates while keeping fine levels known has no effect on behaviour change, as the coefficients of UNKNOWN PROBABILITY KNOWN FINE in each of the two worker regressions are not statistically significantly different from zero.

5.29 For the student sample, we see that not disclosing the fine rate has no significant effect on the likelihood of full compliance. However, non-disclosure of the audit rate results in a significantly higher fraction of income being reported by those who do not fully comply (11 percentage points)[21] but it has no effect on the likelihood of full compliance.

5.30 The evidence from Table 5 highlights an advantage of modelling compliance behaviour using the two-part model. By modelling the extreme of the compliance distribution separately from the remainder, we are able to detect economically meaningful changes in behaviour which would be "averaged out" had we used more standard modelling strategies (see Appendix C for more details).

5.31 The evidence suggests that non-disclosure of information on the fine rate affects various aspects of behaviour differently. On one hand, not knowing the fine rate for under-reporting may increase the fraction of taxpayers who fully comply, while not changing the behaviour of those who evade to a significant degree. On the other hand, not knowing the audit rate seems to increase the compliance level for those students who evade, while not changing their likelihood of full compliance.

5.32 We complete the analysis of Table 5 by discussing UNKNOWN PROBABILITY UNKNOWN FINE FOLLOW-UP. This extra treatment is primarily concerned with the interaction of non-disclosure of information with a follow-up rule: those caught in a given period of the game under-reporting their income would be audited with certainty for the next two periods.

[20] This relationship is significant at the 1% level.
[21] p=0.012.

Table 5: Marginal Effects of Information about Fines/Audit Rages on Compliance

Model	Evade	Percentage of Compliance	Evade	Percentage of Compliance
	Student		Worker	
KNOWN PROBABILITY	-0.04	-0.03	-0.16***	-0.08
UNKNOWN FINE	(0.07)	(0.04)	(0.05)	(0.06)
UNKNOWN	-0.05	0.11**	-0.07	-0.03
PROBABILITY KNOWN FINE	(0.07)	(0.04)	(0.06)	(0.07)
UNKNOWN	-0.15**	0.01	-0.07	-0.03
PROBABILITY	(0.06)	(0.04)	(0.05)	(0.06)
UNKNOWN FINE				
FOLLOW-UP				
Baseline	0.48	0.31	0.22	0.39
N	5,609	2,712	5,949	1,258

Group-level clustered standard errors in parenthesis.
***, **: significant at the 1% and 5% level

5.33 The effect of this treatment on compliance behaviour is not statistically significant for workers. For students, the follow-up treatment is associated with a reduction in the likelihood of evasion (by 15 percentage points). However, the evidence suggests that follow-up audits do not influence the behaviour of those who have chosen to under-declare their income.

Table 6: Marginal Effects of Changes on Fine Levels and Audit Rates on Compliance

Model	Evade	Percentage of Compliance	Evade	Percentage of Compliance
	Student		Worker	
HIGH PROBABILITY	0.06	-0.01	0.08	0.03
LOW FINE	(0.07)	(0.05)	(0.08)	(0.06)
LOW PROBABILITY	-0.12**	0.08	0.11	0.14**
HIGH FINE	(0.06)	(0.09)	(0.07)	(0.06)
HIGH PROBABILITY	-0.20***	0.09	-0.11**	0.08
HIGH FINE	(0.05)	(0.07)	(0.05)	(0.10)
Baseline	0.43	0.36	0.27	0.43
N	3,244	1,412	3,293	903

Group-level clustered standard errors in parenthesis.
***, **: significant at the 1% and 5% level

5.35 Once again, workers respond to the treatment in a very different way to students. Relative to LOW PROBABILITY LOW FINE, increasing the audit probability has no statistically significant impact on behaviour for either students or workers. Raising the penalty for under-reporting while keeping the probability constant has a significant impact on workers who under-report, by making them report 14 percentage points more income, while not significantly impacting their likelihood of evading. The response by students to a high fine is to reduce the likelihood of evasion while not changing the proportion of income reported.

5.36 Finally, and perhaps surprisingly, when we raise both fine and audit rates, we see a similar response by students and workers, in that the rate of evasion goes down for both samples, while the proportion of income reported when evading does not statistically change.

5.37 The difference in response rates between students and workers could be due to the differences in baseline rates of tax evasion. The frequency of full compliance is

much higher among workers than students, while the fraction of income reported when not fully complying is somewhat similar between the two samples. As such, it would be natural to expect that changes in fine rates would lead to students being less likely to evade.

Behavioural Factors Affecting Compliance

5.38 So far our analysis has ignored three potentially important factors which may influence behaviour:

- The possibility that certain communication conditions may be more effective when some information is undisclosed;
- Session-specific events, and payoff-relevant information (e.g. comparing behaviour of those audited to those never audited.)
- Individual-specific behavioural information (e.g. risk aversion, personality measures.)

5.39 We conclude the present section by incorporating all these elements into the econometric analysis.[22] The results from the extended analysis broadly confirm the previous analysis. Compliance behaviour is mostly affected by non-disclosure of information rather than by communication. The regression outputs from this analysis are provided in Appendix C.

5.40 When we include payoff-relevant information into the analysis, our findings are broadly applicable to students and workers. Being able to solve an extra slider in the 'real effort task' (which equates to one more ECU in taxable income) results in a 0.5/0.6 percentage point increase in declared income by workers/students. In other words, more able participants in a given period will declare a higher fraction of their taxable income compared to low ability participants. However, we find that as the experiment progresses, there is a very small trend towards non-compliance – while statistically significant, the increase is very small and economically meaningless.

5.41 We find that being audited but not caught in a given period significantly reduces the likelihood of evasion in the subsequent period, relative to those who were not audited (12 percentage points for students, 7 percentage points for workers).

5.42 Interestingly, and perhaps counter-intuitively, those who were audited *and caught* were more likely to evade in the following period than those who were audited and *not caught* because they had fully declared their income.

5.43 Similarly, participants who were audited in the first five periods of the experiment were less likely to evade than those who were not audited early on; however those who were audited *and caught* in the first five periods were more likely to evade than those audited in the first five periods but *not caught*.

5.44 This likely reflects two factors at play:

- Individuals who under-report today are more likely to under-report tomorrow, regardless of whether they were audited;
- Participants may have an expectation of earnings (or an earnings target) in the experiment, and paying a fine may lead to more risk-seeking behaviour in order to achieve that target.

[22] We also controlled for different group sizes as discussed in Section 3.

We are unable to distinguish between the two hypotheses, but we note that similar motivations may drive real behaviour in the field.[23]

5.45 We find risk aversion is not a significant predictor of tax compliance in our experiment. We can interpret this as evidence that even though stakes were relatively high, they were not sufficiently high for risk aversion to play a role.[24]

5.46 We also find very limited support for differences in personality measures in the Big Five model. We find a small positive marginal effect of Emotional Stability on the likelihood to evade for students. For the worker sample, there is no evidence of association between the personality measures and compliance behaviour.[25]

[23] Spicer and Hero (1985) and Webley (1987) find that participants in their experiments who were audited report more income after their audits than participants who were never audited.
[24] The role of risk attitudes in explaining behaviour in the lab has long been the subject of theoretical and empirical analysis. Rabin (2000) showed that if experimental subjects were to reject low-stake gambles it would imply absurdly high coefficients of risk aversion, as illustrated by this example. "Suppose that, from any initial wealth level, a person turns down gambles where she loses $100 or gains $110, each with 50% probability. Then, she will turn down 50-50 bets of losing $1,000 or gaining *any* sum of money." (p.1282).
[25] All coefficients are essentially equal to zero.

6. Discussion

6.1 This chapter will discuss the interpretation of the experimental results and suggest some implications of the results for HMRC. It will also consider the role and conduct of future experiments on tax compliance.

Non-Student Behaviour

6.2 The experiment has delivered weak evidence that communication leads to a higher rate of compliance among the worker subject pool. This is a surprising result if a traditional view is taken of the compliance decision. The taxpayers need to know the probability of audit and the rate of fine in order to make the correct decision for the compliance 'gamble'. Any credible information of these variables should be incorporated into the decision problem and the choice should change. That this does not happen in the experiment shows either that the communication is not credible, that the information is credible but does not add to existing knowledge, or that the traditional perspective is incorrect.

6.3 Our inference is that the explanation is a combination of the first and third. The justification for this is that communication has more impact (though is still not statistically significant) when the number of subjects involved in the communication is restricted. This conforms to the idea of a small network with trust and credible information. Furthermore, the non-traditional behavioural perspective accepts that the decision is made on the basis of a subjective probability that is individual-specific and need bear only a very weak relation to the true objective probability. Since the subjective probability is a psychological construct it need not be affected by communication of information about the objective probability.

6.4 Similar arguments can be advanced to explain why awareness of the audit rate does not modify compliance behaviour to any great extent. For example, if the psychological perspective of each subject is to believe that they are targeted by the HMRC ("I know 10% of people are audited, but they are sure to audit me.") then knowing the audit rate will not shift the psychological perspective. This is counter to the traditional view but consistent with a behavioural model.

6.5 A lack of knowledge of the fine increases the proportion of workers who are fully compliant. This suggests that the average subject believes the fine to be higher than it is in the experimental setting. Since the fine rates used in the experiment are in line with those in practice, this implies an excessive average belief about the fine rate.

6.6 The combined approach of increasing the audit rate and the fine rate was much more effective than increasing either instrument independently. The increase in compliance achieved by the combination was statistically significant and large. This must reflect the fact that an effective policy should be credible: there is no benefit of point promising more audits if this is not backed by an enhanced punishment.

6.7 Interestingly, the effect of being audited and caught under-reporting in the results in a higher likelihood of evading later on. This could be due to the fact that people who are caught are those more likely to evade in the first place. Alternatively, if participants have an earnings target, the reduction in income following the fine makes participants more risk seeking than before in the attempt to reach this earnings target.

6.8 Note that either motivation is likely to apply to real taxpayers. Some individuals may be more prone to evading tax; other individuals may have expectations about their lifetime earnings, and such expectations could inform their tax compliance behaviour. Our finding suggests that increasing the likelihood (potentially up to certainty) of auditing taxpayers who have previously not complied for several years after they were caught, and making this known among taxpayers, may be an effective policy to be pursued by HMRC.

Differences between Students and Workers

6.9 The original feature of the experiment was the use of a subject pool with (approximately) equal numbers of students and workers. Most previous experiments on tax compliance have used student subjects only. The use of a mixed subject pool permits comparison of the compliance behaviour of the two groups and an assessment of whether a student subject pool is appropriate. It also provides further insight into the factors that determine the compliance decision.

6.10 We observe that students were more responsive than workers to changes in parameters in the experiment, such as the level of fines for non-compliers, or probability of audit.[26] This could suggest, as pointed out by a reviewer, that advertising audit rates may be a counter-productive measure: while it has no effect on the compliant sub-sample (workers), it has a detrimental effect on the sub-sample which is more responsive to economic incentives (students). Unlike past evidence, we find the audit rate elasticity is close to zero, while the fine elasticity is slightly above 0.1. In their survey of the experimental literature of tax compliance, Alm and Jacobson (2007) found reported income-audit rate elasticities in the range of 0.1-0.2, and reported income-fine rates of less than 0.1.

6.11 The results clearly show that workers have a higher compliance rate in the experiment than students and that students are more responsive to the incentives than workers. In addition, workers respond to changes in information about the fine rate while students respond to changes in information about the audit rate.

6.12 Our evidence is in contrast to recent experimental work on tax compliance by Alm et al (2010), which compares the behaviour of university students to university staff. That study found students and staff differed in terms of the compliance level, but not in the way their responded to changes in audit probability.

6.12 Our evidence is a cautionary note on studying tax compliance using non-taxpayers as a subject pool. While it is premature to draw definitive conclusions about how what subjects one should use in experimental research in this area, it is important to further understand why there are differences in behaviour between the two samples.

6.13 Potential reasons for the difference include the different recruitment processes for the two different samples; the fact that most of our worker sample took part in the experiment online, as opposed to the lab and different show-up fees and incentives (although these reflect the different opportunity costs of time for each sample). Students may also have been less sensitive than workers to the explicit tax framing used in the instructions.[27] It is beyond the scope of this report to explore

[26] This is supported by the econometric evidence in Section 5, which shows students' likelihood of non-compliance and/or the degree of compliance significantly changing with some treatments, while observing no such responsiveness in the worker sample, and vice-versa.
[27] We thank our independent reviewers for pointing these issues.

these explanations for the differences in behaviour between the two subject pools, and we leave them for future study.

6.14 The finding of a difference in behaviour between the two subject pools is not consistent with the traditional model of tax compliance. If that model were true then the two groups would have the same behaviour.

6.15 Our interpretation of the results is that the student subject pool have had little or no socialisation in the process of tax payment. They therefore enter the experiment with the aim is to extract the maximum payoff. Hence, they essentially follow a strategy to maximise the quantity of money taken from the experiment. In each case that there is a reaction of the student pool to a treatment it is consistent with this perspective.

6.16 In contrast the subject pool of workers have been socialised into the tax payment process. As a result they ought to carry into the experiment the social customs, beliefs, and norms of behaviour that they have adopted outside the laboratory. The behaviour they exhibit is therefore consistent not with the treatments inside the experiment but with their perception of how to behave outside the laboratory. The behaviour is consistent with a behavioural model in which most people follow the social custom of full compliance and hold subjective beliefs about the probability of audit. Behaviour in the laboratory is then insensitive to treatments because behaviour is driven by attitudes and beliefs that have been established over a long period of socialisation outside the laboratory.

6.17 An alternative perspective on this finding, pointed out by one of our independent reviewers is that our worker sample, who pay their income tax through PAYE, may be attitudinally compliant, given that non-compliance is not possible in their daily experience. Students, in contrast may approximate the behaviour of taxpayers who are not attitudinally compliant, and as such may provide useful and policy-relevant insights into tax compliance. Recruiting self-employed participants, though more difficult, would help understand this issue.

6.18 The results also suggest that future experiments on tax compliance need to consider a wider range of treatments than have so far been employed. This could be undertaken in a laboratory setting (a term which encompasses online participation in experiment). In particular, further research is warranted to understand the role communication among participants play (if any) in determining compliance levels. The relative lack of effect of communication on compliance may be due to the fact that audit rates are relatively easy to infer, and that knowledge is established after a few rounds. It may also be of limited use to participants given the relative simplicity of the experiment. A potential extension of the current setup could involve different groups with different audit rates.

6.19 One could also increase the complexity of the tax return task and introduce intermediaries, which could proxy the role of tax preparers. Recent evidence (Hasseldine et al. 2005) suggests the role of communication between a tax authority and taxpayers is more effective with taxpayers who do not resort to tax accountants. This effect could result from self-selection (those who self-report may be more prone to evasion), or it could be due to opportunity cost of time – in the absence of a warning, taxpayers who self-report do not spend enough time understanding the intricacies of the filling process. An experimental investigation could be a means to understand which factor is at play.

6.20 In short, laboratory experiments should be seen as an important complement to analyse taxpayer behaviour in tandem with more traditional empirical methodologies, such as randomised controlled experiments on tax compliance (e.g. Kleven et al. 2011). The focus should be on the behavioural mechanisms that explain observed non-compliance from field experiments.

References

Alalehto, T. (2003) “Economic crime: Does personality matter?” *International Journal of Offender Therapy and Comparative Criminology,* 47(3), 335-355.

Allingham M. and Sandmo, A. (1972) “Income tax evasion: a theoretical analysis.” *Journal of Public Economics* 1, 323-338.

Alm, J., Bloomquist, K.M., and McKee, M. (2010) “On the External Validity of Tax Compliance Experiments”, *mimeo*.

Alm, J., Cherry, T., and McKee, M., and Jones, M.L. (2009) "Encouraging filing via tax credtis and social safety nets", The IRS Research Bulletin, Number 1500.

Alm, J., Cherry, T., and McKee, M., and Jones, M.L. (2010) "Investigating behavioral responses to positive inducements for filing tax returns", Working paper 10-11, Appalachian State University.

Alm, J. and Jacobson, S. (2007) “Using Laboratory Experiments in Public Economics” *National Tax Journal*, 60(1), pp. 129-152.

Andreoni, J., Erard, B., and Feinstein, J. (1998) “Tax Compliance.” *Journal of Economic Literature*, 36, 818-860.

Barham, E. and Fox, K. (2011) “Compliance Perceptions Survey: Individuals 2008-10” HM Revenue and Customs Research Report 156.

Cleave, B.L., Nikiforakis, N. and Slonim, R. (2012) “Is There Selection Bias in Laboratory Experiments? The Case of Social and Risk Preferences.” *Experimental Economics*, forthcoming.

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

Costa, P. T., Jr., and McCrae, R.R. 1992. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources

Crawford, V. (1998) “A Survey of Experiments on Communication via Cheap Talk.” *Journal of Economic Theory* 78, 286-298.

Fischbacher, Urs. (2007) “z-Tree - Zurich toolbox for Readymade Economic Experiments.” *Experimental Economics* 10(2): 171-178.

Fonseca, M. and Myles, G.D. (2012) “Experiments on tax compliance: a Survey.” *mimeo*, University of Exeter.

Gill, D., and Prowse, V. (2012) “A Structural Analysis of Disappointment Aversion in a Real Effort Competition.” *American Economic Review* 102(1): 469-503.

Greiner, B. (2004) “An Online Recruitment System for Economics Experiments.” *University Library of Munich, Germany, Mpra Paper*.

Hashimzade, N., Myles, G.D. and Tran-Nam, B. (2012) “Applications of behavioural economics to tax evasion.” *Journal of Economic Surveys (forthcoming).*

Hasseldine, J., Hite, P., James, S. and Toumi, M. (2005) “Carrots, Sticks, Sole Proprietors and Tax Accountants” *The IRS Bulletin: Recent Research on Tax Administration and Compliance*, pp. 191-210.

HMRC (2012) *Measuring Tax Gaps 2012*, London.

Holt, C.A., and Laury, S.K. (2002) “Risk Aversion and Incentive Effects.” *American Economic Review* 92, 1644-1655.

John, O.P., and Srivastava, S. (1999) “The Big Five trait taxonomy: History, measurement and theoretical perspectives.” In L.A. Pervin, and O.P. John (eds.) *Handbook of personality theory and research*, pp. 102-138. New York: Guildford.

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

McCrea, R.R. and Costa, P.T., Jr. (1999) “A five-factor theory of personality.” In L.A. Pervin, and O.P. John (eds.) *Handbook of personality theory and research*, 139-153. New York: Guildford.

Myles, G.D. and Naylor, R.A. (1996) “A model of tax evasion with group conformity and social customs.” *European Journal of Political Economy*, 12, 49-66.

Pyle, D.J. (1991) “The economics of taxpayer compliance.” *Journal of Economic Surveys* 5, 163-198.

Rabin, M. (2000) “Risk aversion and expected-utility theory: A calibration theorem.” *Econometrica* 68, 1281-1292.

Sandmo, A. (2005) “The theory of tax evasion: A retrospective view.” *National Tax Journal* 58, 643-663.

Slemrod, J. and Yitzhaki, S. (2002) “Tax avoidance, evasion and administration.” in A.J. Auerbach, A.J. and M. Feldstein (eds.) *Handbook of Public Economics*, Volume 3, Amsterdam: North Holland.

Spicer, M.W. and Hero, R.E. (1985) “Tax Evasion and Heuristics: A Research Note” *Journal of Public Economics*, 26(2), pp. 263-267.

Thaler, R.H. (1994) *The Winner's Curse: Paradoxes and Anomalies of Economic Life*. Princeton: Princeton University Press.

Webley, P. (1987) “Audit Probabilities and Tax Evasion in a Business Simulation” *Economics Letters*, 25(3), pp. 267-270.

Woods, S. A., and Hampson, S. E. (2005) “Measuring the Big Five with single items using a bipolar response scale.” *European Journal of Personality* 19(5), 373-390.

Yitzhaki, S. 1974. “A note on income tax evasion: A theoretical analysis.” *Journal of Public Economics* 3, 201-202.

Appendix A: Experimental Materials

Sample Instruction Sets (FREE UNKNOWN PROBABILITY UNKNOWN FINE FOLLOW UP, NON-STUDENTS)

Instruction set

Welcome to our experiment. You will have **10 minutes** in which to read these instructions; please read them carefully because you will **NOT** be able to refer back to them once the 10 minutes is up. Your payoff in this experiment will depend on what decisions you take. It is therefore important that you understand the rules of the experiment.

This experiment will be divided into 2 parts: part A and part B. We will now explain how part A will work. Once part A is over, we will show instructions for part B.

In this experiment, your payoff will be in Experimental Currency Units (ECU). 15 ECU are worth £1**.** After the experiment is over, we will convert the sum of your payoff for the whole session into pounds and pay you via a bank transfer. In addition to your earnings during the session, you will receive £20 for participating.

Part A

In this part of the experiment, you will be playing with a group of people. This group will consist of 5 people including you. The four other participants will be connected to the same session as you through the software that you are running in your computer. Each player will be given an identifying letter: A, B, C, D and E. you will see your ID on the screen.

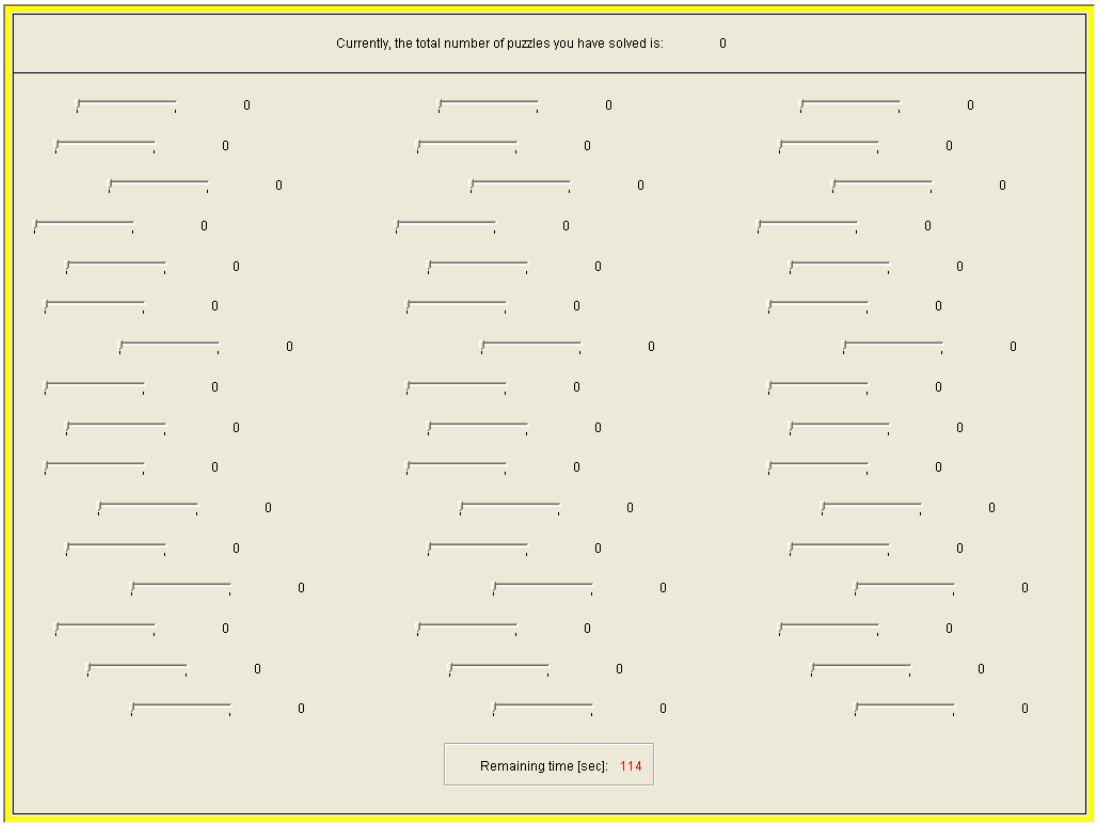
This part of the experiment is divided into 15 rounds.

In each round, you will have the opportunity to make money by performing a task, which will last **120 seconds**.

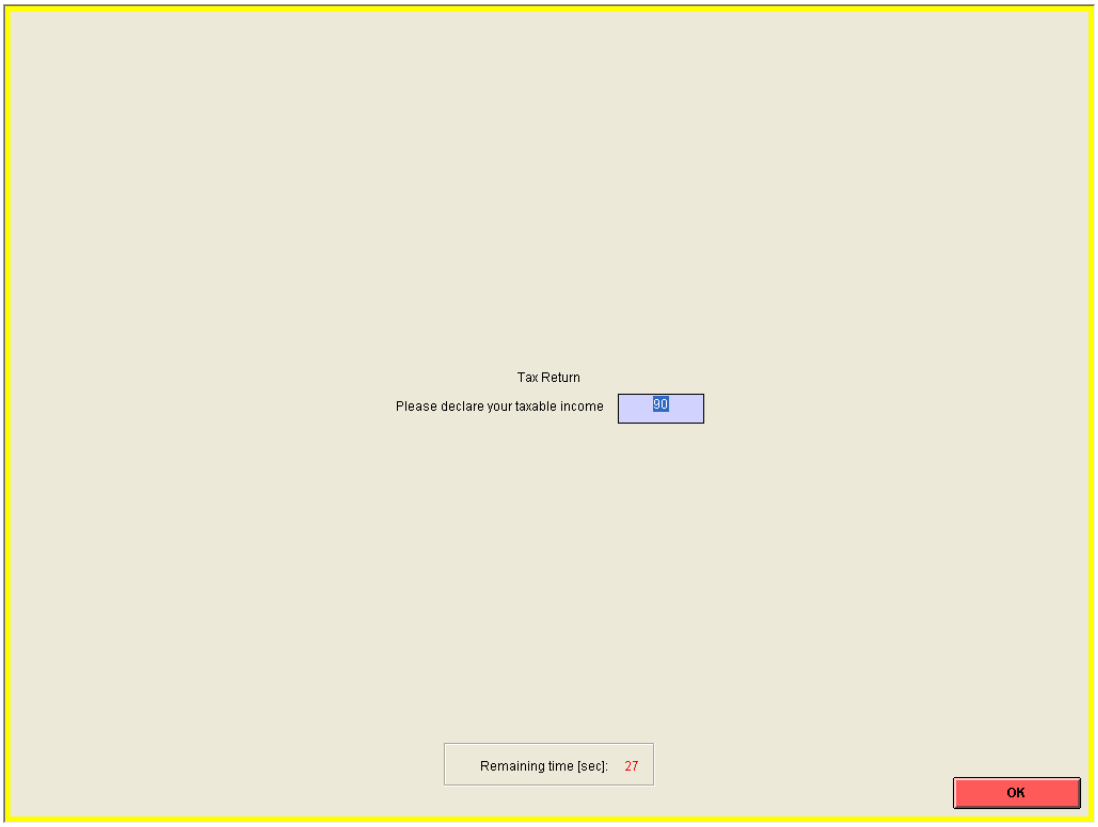
The task consists of a screen with 48 sliders. Each slider is initially positioned at 0 (the far left of the line) and can be moved as far as 100 (the far right of the line). Each slider has a number to its right, which tells you its current position. You use the mouse to move the slider. You do this by dragging the slider along the line. You can change the position of each slider as many times as you wish.

You solve each slider by placing it at 50. For each slider you solve, you will receive 1 ECU.

See below for a preview of what the screen looks like.



Once you have finished your task, the computer will tell you how much income you earned. Your income is taxable, so you will then have to fill a tax return. The tax rate on your income is 25%. This means that for every 10 ECU you earn, you must pay 2.5 ECU in tax. The figure below shows you the tax return you must fill.



Once you complete your tax form, the tax authority will process all forms. The fraction of people audited in every period is unknown.

If you are not audited, then your payoff for the round will be equal to the amount of income you made in the round minus the tax you paid on the income you reported on your tax form.

If you are audited and if you reported your income accurately, then nothing further will happen; your payoff for the round will be that same as if you had not been audited.

If you are audited and if you reported less income than you actually earned, then you will have to pay the extra amount of tax due to the authority. In addition you will pay a fine for each ECU you did not report in your tax form. The value of the fine is unknown. However, in this case the tax authority will audit you with certainty for the two periods after you pay the fine.

If you are audited and if you reported more income than you actually earned, then you will receive a refund of the tax you over-paid.

You will then see a screen which summarises what has happened:

- How much income you earned by solving sliders
- The amount of income you reported to the tax authority
- Whether you were audited or not by the tax authority
- Your final payoff for the round.

You will have an OK button on the bottom right-hand corner of the screen. Clicking that button takes you to the following screen.

In the following screen, you will be able to communicate with the other participants in the experiment using a chat box for 30 seconds. You may communicate about whatever you like. The only restrictions are that you may not use offensive language and you may not identify yourself (i.e. you may not reveal your name.) The following picture shows you how this chat box will look like.

Once the 30 seconds run out, the chat box will close and the round will be over.

Summary

In short this part of the experiment will consist of five stages:

STAGE 1: You earn income by operating slides.

STAGE 2: You declare your income to the tax authority

STAGE 3: The tax authority audits a fraction of the population

STAGE 4: You get your final payoffs, which depend on how your report your income and whether you were audited or not.

STAGE 5: You can communicate with people in the experiment.

Appendix B: Recruitment Materials

B1 Non-Student Invitation Letter

Dear «FirstName» «LastName»
{Registration ID: «RegistrationID»}

I am writing to advise you about a project for which Saros is currently recruiting; based upon the information you gave us when you registered with Saros you may be able to take part.

The project is a different method of market research and involves taking part in an online problem solving exercise about financial decision making. You will be required to take part in an online exercise and you will be logged on anonymously at the same time as 4 other people. The exercise will take no more than 90 minutes to complete and you can earn between £20 and £60 depending on your decisions in the exercise. You will be required to download some software in order to take part in the session and you must complete the exercise in one sitting. The sessions will be taking place *(Dates)* at a time convenient to you.

General eligibility is as follows:

- You must be aged between 18 and 65

If this applies to you and you would like to register your interest in participating please click the link below, to answer a few quick questions.

[url here]

Please note that this webform is a 2-minute screening questionnaire, which we will use to determine whether or not you are part of the targeted audience our client wishes to consult. We shall endeavour to let you know as soon as possible whether or not you have been selected. Thank you for your patience.

With kind regards,

B2 FEELE Consent Form

In order to participate in experiments conducted in the Feele lab at the University of Exeter Business School, you must agree with the rules of our lab and our privacy policy. By registering for Feele lab experiments, you indicate your availability to participate in economic experiments conducted by researchers from the University of Exeter.

Please read the following points very carefully.

- For each experiment, a certain number of registered persons will receive an invitation email. Only invited persons have the opportunity to take part in the experiment.
- If you receive an email inviting you to participate in a particular experiment, you are not obliged to respond: participation is entirely voluntary. If you do wish to participate, you may follow the instructions in the email and book one session of the experiment. By booking a session, you are making a commitment to show up ON TIME to that session. If you arrive two or three minutes late, you will NOT be considered to have arrived 'on time'.

- If you do participate in the experiment, your payment will depend on an element of luck as well as on the decisions made both by you and by other participants with whom you are grouped. You will be paid a show-up fee in addition to your payment from the experiment: this will be stated clearly in the instructions for the experiment.
- Once the session is under way, your continuing participation is voluntary and you have the right to withdraw at any time. If you do withdraw, however, or fail to complete the session for whatever reason, you will only receive your show-up fee.
- If we are forced to abandon a session, e.g. due to computer failure, all participants will be paid a show-up fee whose magnitude depends on the length of time the abandoned session had been running.
- All participants will be expected to comply with the rules of the experiment as defined in the instructions. If you do not understand the instructions, then the responsibility is yours to seek clarification.
- Please remember that if you book a particular session of an experiment and then do not take part or if you are late, that the session may have to be abandoned because of the need to have an exact number of participants. This inconveniences a number of would-be participants who have made the effort to show up on time. It also costs us money and we will hold you liable for such costs if you do not give us advance notice of at least 24 hours.

I have read the above points and agree with the following statements:

- The registration information which I have provided is accurate and I have not attempted to register more than once.
- I have been informed that if I do not turn up or turn up late for a session which I have booked, then this may lead to the session having to be abandoned. I am aware that I can be made liable for costs incurred from abandoned sessions if I do not inform the Feele lab at least 24 hours ahead of time that I will not be able to attend and do not give a genuine explanation as to why this is. I also understand that I will need a very compelling reason, i.e. accident or sudden serious illness, for failing to give such advance notice.

Privacy policy

Personal data

- We will collect your bank account as part of the recruitment to take part in scientific experiments. We do not disclose this data to third parties.
- We use the data to invite a scientifically-chosen subset of subscribers to register for our experiments
- There is no link between the computer system recording the experiment data and our recruitment system, both in terms of your personal data, as well as your bank details.)

Experiment data

- We use a computer system to record the decisions made by participants in our experiments.
- The experimenters analyse this data for the purposes of scientific research. The decision data cannot be used to identify individual participants and in this sense participation in our experiments is anonymous.
- The anonymous post-analysis data will be used to produce academic research and to deliver lectures. This work will be published.

Do you agree with the rules and the privacy policy?

B3 Debrief Document

Dear Participant,

Thank you for participating in the Feele study. We would also like to briefly explain the purpose of this study to you.

This study was designed to understand the determinants of taxpayer compliance. This was accomplished by asking participants to perform the slider task, which determined your pre-tax payoff. We then measured tax compliance by calculating the fraction of your pre-tax payoff you declared.

The purpose of this study was to see what effect allowing people to communicate has on tax compliance. To do this we varied the rules of the experiment from session to session. By doing this, we hope to gain a better understanding of taxpayer behaviour. In some sessions, participants knew the fine they would pay if they under-declared income and were audited; in other sessions they did not. In some sessions we also informed participants of how frequently the computer would audit them; in other sessions we did not. Finally, we also allowed participants to communicate with each other in some sessions; in other sessions we did not.

This research is funded by HMRC. We did not state this when we invited you to participate because we did not want to bias your expectations about the study before you participated.

We would like to reassure you that your data is fully anonymous, which means it is impossible to link your responses in the experiment with your identity. Researchers at the University of Exeter follow strict ethics standards, which include protecting the privacy of our participants.

The HMRC will not have access to your personal data; they will only be given the data from the experiment, which will not have any names or other identifying information.

We plan to use the results from the study in reports and scientific journal articles. The results in these reports will be presented in aggregated form – for example the average compliance rate across all participants in our study.

If you nevertheless wish to opt out from this study, we will delete your data from the project. Opting out will not affect your payment from the experiment. To opt out from the study, please print and sign your name below and post this document to the following address: FEELE Lab, University of Exeter, Streatham Court Room 0.37, Rennes Drive, Exeter EX4 4PU.

FULL NAME:

ADDRESS:

SIGNATURE:

Appendix C: Regression Outputs

As discussed in the main report, estimating the effect of our treatments using ordinary least squares (OLS) will ignore important features of this distribution. It will also not guarantee that the predicted values obtained from the regression analysis will lie in the interval between zero and one. In other words, due to the nature of this type of analysis, the results may suggest that some individuals declare less than 0% of their income, or that others declare more than 100% of their income.

A commonly used econometric specification when studying fractional data of this form is the Tobit model. However, the Tobit model is appropriate to model data that is *censored*, either in a given interval or by a particular lower or upper bound. The classic example is estimating the demand for cinema tickets while only observing data on ticket sales which has an upper bound equal to the total number of seats. However, it is difficult to justify the use of the Tobit model for data that is *defined* only for a given interval, which is the case with our data set.

Table 7: Worker evasion decision regression estimates

	Evade 1		Evade 2		Evade 3	
	Coef	S.E	Coef	S.E	Coef	S.E
Free x Unknown Prob Unknown Fine - Follow Up	-0.22	(0.23)	-0.00	(0.26)	0.12	(0.31)
Free x Unknown Prob Known Fine	-0.24	(0.28)	0.01	(0.26)	0.06	(0.24)
Free x Known Prob Unknown Fine	-0.83***	(0.28)	-0.57**	(0.26)	-0.81**	(0.36)
Restrict x Unknown Prob Unknown Fine - Follow Up	-0.33	(0.22)	-0.20	(0.23)	-0.26	(0.25)
Restrict x Unknown Prob Known Fine	-0.15	(0.29)	-0.28	(0.28)	-0.61	(0.31)
Restrict x Known Prob Unknown Fine	-0.86***	(0.21)	-0.48	(0.25)	-0.48	(0.28)
Public x Unknown Prob Unknown Fine - Follow Up	-0.18	(0.28)	-0.15	(0.27)	-0.26	(0.26)
Public x Unknown Prob Known Fine	-0.15	(0.29)	0.10	(0.30)	0.15	(0.30)
Public x Known Prob Unknown Fine	-0.86***	(0.21)	-0.60***	(0.19)	-0.54***	(0.20)
NoComm x Unknown Prob Unknown Fine - Follow Up	-0.36	(0.25)	-0.18	(0.26)	-0.16	(0.33)
NoComm x Unknown Prob Known Fine	-0.15	(0.29)	0.09	(0.28)	0.08	(0.26)
NoComm x Known Prob Unknown Fine	-0.30	(0.29)	-0.22	(0.25)	-0.40	(0.23)
NoComm x High Prob High Fine	-0.56**	(0.24)	-0.36	(0.22)	-0.41	(0.23)
NoComm x High Prob Low Fine	0.05	(0.26)	0.09	(0.24)	0.12	(0.29)
NoComm x Low Prob High Fine	0.13	(0.24)	0.32	(0.22)	0.53	(0.28)
Ability			0.00	(0.01)	-0.01	(0.01)
Payment			0.00	(0.00)	0.0002***	(0.00005)
Audited & Not Caught (t-1)			-0.36***	(0.06)	-0.28***	(0.06)
Audited & Caught (t-1)			1.66***	(0.13)	1.38***	(0.14)
Early Audited & Not Caught			-2.93***	(0.22)	-1.95***	(0.13)
Early Audited & Caught			5.87***	(0.22)	4.93***	(0.18)
Worker x Lab			0.12	(0.11)	0.33**	(0.15)
Group Size			-0.05	(0.07)	-0.07	(0.08)
Extraversion					0.03	(0.03)
Agreeableness					-0.05	(0.03)
Emotional Stability					0.04	(0.03)
Conscientiousness					0.01	(0.03)
Openness					0.01	(0.04)
Risk Aversion					0.03	(0.02)
Personal Income					0.08	(0.06)
Household Income					-0.04	(0.04)
Age					-0.01	(0.01)
Male					0.02	(0.14)
Work Experience					-0.01	(0.01)
Constant	-0.44**	(0.18)	-0.81**	(0.32)	-0.78	(0.57)
N	7,446		6,959		5,435	

Group-level clustered standard errors in parenthesis. ***, **: significant at the 1% and 5% level

Table 8: Non-compliant worker percentage of income reported regression estimates

	% of Compliance 1		% of Compliance 2		% of Compliance 3	
	Coef	S.E	Coef	S.E	Coef	S.E
Free x Unknown Prob Unknown Fine - Follow Up	0.11	(0.23)	0.18	(0.26)	0.55	(0.38)
Free x Unknown Prob Known Fine	-0.20	(0.29)	-0.15	(0.28)	0.10	(0.33)
Free x Known Prob Unknown Fine	-0.36	(0.26)	-0.35	(0.26)	0.01	(0.45)
Restrict x Unknown Prob Unknown Fine - Follow Up	-0.25	(0.18)	-0.23	(0.20)	-0.16	(0.29)
Restrict x Unknown Prob Known Fine	-0.02	(0.28)	0.08	(0.29)	0.16	(0.36)
Restrict x Known Prob Unknown Fine	0.10	(0.27)	0.21	(0.30)	0.20	(0.43)
Public x Unknown Prob Unknown Fine - Follow Up	-0.09	(0.22)	-0.04	(0.24)	0.41	(0.33)
Public x Unknown Prob Known Fine	-0.07	(0.25)	-0.04	(0.28)	0.18	(0.34)
Public x Known Prob Unknown Fine	-0.47	(0.26)	-0.42	(0.25)	-0.27	(0.31)
NoComm x Unknown Prob Unknown Fine - Follow Up	-0.18	(0.22)	-0.15	(0.24)	0.22	(0.31)
NoComm x Unknown Prob Known Fine	-0.03	(0.28)	0.03	(0.29)	0.41	(0.38)
NoComm x Known Prob Unknown Fine	-0.26	(0.23)	-0.25	(0.25)	0.01	(0.32)
NoComm x High Prob High Fine	0.08	(0.26)	0.23	(0.29)	0.55	(0.32)
NoComm x High Prob Low Fine	-0.03	(0.18)	-0.01	(0.20)	0.12	(0.29)
NoComm x Low Prob High Fine	0.25	(0.18)	0.29	(0.20)	0.23	(0.32)
Ability			0.01	(0.01)	0.01**	(0.006)
Payment			-0.0002***	(0.00005)	-0.0002***	(0.00006)
Audited & Not Caught (t-1)			0.04	(0.08)	0.10	(0.09)
Audited & Caught (t-1)			-0.09	(0.08)	-0.22**	(0.10)
Early Audited & Not Caught			0.04	(0.06)	0.02	(0.08)
Early Audited & Caught						
Worker x Lab			0.09	(0.12)	0.13	(0.17)
Group Size			-0.02	(0.06)	-0.09	(0.07)
Extraversion					-0.01	(0.04)
Agreeableness					0.02	(0.03)
Emotional Stability					0.04	(0.03)
Conscientiousness					0.05	(0.04)
Openness					-0.05	(0.03)
Risk Aversion					0.01	(0.04)
Personal Income					-0.06	(0.06)
Household Income					-0.01	(0.04)
Age					-0.33**	(0.15)
Male					0.02	(0.01)
Work Experience					0.02	(0.01)
Constant	-0.19	(0.12)	0.10	(0.25)	0.78	(0.72)
N	1,720		1,588		1,220	

Group-level clustered standard errors in parenthesis. ***, **: significant at the 1% and 5% level

Table 9: Student evasion decision regression estimates

	Evade 1		Evade 2		Evade 3	
	Coef	S.E	Coef	S.E	Coef	S.E
Free x Unknown Prob Unknown Fine - Follow Up	-0.32	(0.18)	-0.20	(0.21)	-0.59	(0.42)
Free x Unknown Prob Known Fine	0.01	(0.31)	0.04	(0.34)	0.11	(0.37)
Free x Known Prob Unknown Fine	0.08	(0.25)	0.12	(0.28)	0.36	(0.34)
Restrict x Unknown Prob Unknown Fine - Follow Up	-0.50***	(0.19)	-0.48**	(0.20)	-0.57***	(0.19)
Restrict x Unknown Prob Known Fine	-0.05	(0.25)	-0.03	(0.25)	-0.02	(0.24)
Restrict x Known Prob Unknown Fine	-0.14	(0.17)	-0.11	(0.19)	-0.01	(0.20)
Public x Unknown Prob Unknown Fine - Follow Up	-0.21	(0.19)	-0.15	(0.10)	-0.09	(0.21)
Public x Unknown Prob Known Fine	-0.20	(0.23)	-0.14	(0.24)	-0.31	(0.25)
Public x Known Prob Unknown Fine	-0.30	(0.24)	-0.29	(0.24)	-0.31	(0.26)
NoComm x Unknown Prob Unknown Fine - Follow Up	-0.56**	(0.25)	-0.58**	(0.25)	-0.34	(0.25)
NoComm x Unknown Prob Known Fine	-0.30	(0.24)	-0.31	(0.24)	-0.01	(0.20)
NoComm x Known Prob Unknown Fine	0.00	(0.24)	0.03	(0.25)		
NoComm x High Prob High Fine	-0.75***	(0.19)	-0.72***	(0.21)	-0.68***	(0.25)
NoComm x High Prob Low Fine	-0.06	(0.21)	-0.07	(0.22)	-0.03	(0.23)
NoComm x Low Prob High Fine	-0.53***	(0.19)	-0.52***	(0.20)	-0.44	(0.23)
Ability			-0.01	(0.01)	-0.01**	(0.006)
Payment			0.0005***	(0.0001)	-0.0005***	(0.00008)
Audited & Not Caught (t-1)			-0.33***	(0.07)	-0.32***	(0.08)
Audited & Caught (t-1)			1.08***	(0.10)	1.04***	(0.13)
Early Audited & Not Caught			-2.90***	(0.12)	-3.07***	(0.15)
Early Audited & Caught			6.19***	(0.20)	6.25***	(0.21)
Extraversion					-0.01	(0.02)
Agreeableness					0.01	(0.03)
Emotional Stability					0.09***	(0.03)
Conscientiousness					0.00	(0.02)
Openness					0.04	(0.02)
Risk Aversion					0.02	(0.02)
Age					-0.05	(0.07)
Male					-0.04	(0.10)
Constant	0.15	(0.15)	-0.21	(0.18)	0.12	(1.27)
N	7,125		6,650		4,988	

Group-level clustered standard errors in parenthesis. ***, **: significant at the 1% and 5% level

Table 10: Non-compliant student percentage of income reported regression estimates

	% of Compliance 1		% of Compliance 2		% of Compliance 3	
	Coef	S.E	Coef	S.E	Coef	S.E
Free x Unknown Prob Unknown Fine - Follow Up	-0.12	(0.13)	-0.20	(0.17)	-0.38	(0.45)
Free x Unknown Prob Known Fine	0.11	(0.17)	0.15	(0.16)	-0.00	(0.16)
Free x Known Prob Unknown Fine	-0.14	(0.20)	-0.14	(0.21)	-0.08	(0.26)
Restrict x Unknown Prob Unknown Fine - Follow Up	0.04	(0.26)	-0.02	(0.27)	-0.64***	(0.19)
Restrict x Unknown Prob Known Fine	0.49**	(0.21)	0.51**	(0.20)	0.46	(0.24)
Restrict x Known Prob Unknown Fine	-0.30	(0.17)	-0.30	(0.20)	-0.49***	(0.18)
Public x Unknown Prob Unknown Fine - Follow Up	0.20	(0.15)	0.18	(0.16)	0.07	(0.17)
Public x Unknown Prob Known Fine	0.36**	(0.18)	0.39**	(0.19)	0.10	(0.21)
Public x Known Prob Unknown Fine	-0.26	(0.18)	-0.30	(0.20)	-0.18	(0.26)
NoComm x Unknown Prob Unknown Fine - Follow Up	0.04	(0.20)	-0.00	(0.19)	0.15	(0.21)
NoComm x Unknown Prob Known Fine	0.26	(0.19)	0.25	(0.20)	-0.03	(0.16)
NoComm x Known Prob Unknown Fine	0.33***	(0.12)	0.30**	(0.13)		
NoComm x High Prob High Fine	0.40**	(0.19)	0.36	(0.19)	0.28	(0.24)
NoComm x High Prob Low Fine	0.15	(0.13)	0.23	(0.15)	0.21	(0.16)
NoComm x Low Prob High Fine	0.37	(0.24)	0.37	(0.25)	0.24	(0.30)
Ability			0.01***	(0.005)	0.02***	(0.005)
Payment			-0.0009***	(0.00008)	-0.0009***	(0.00009)
Audited & Not Caught (t-1)			0.10	(0.07)	0.14	(0.09)
Audited & Caught (t-1)			-0.39***	(0.09)	-0.45***	(0.11)
Early Audited & Not Caught					-0.01	(0.08)
Early Audited & Caught						
Extraversion					0.04	(0.03)
Agreeableness					-0.00	(0.03)
Emotional Stability					-0.04	(0.02)
Conscientiousness					-0.00	(0.02)
Openness					0.02	(0.03)
Risk Aversion					0.00	(0.03)
Age					-0.10	(0.08)
Male					-0.37***	(0.12)
Constant	-0.57***	(0.08)	0.02	(0.15)	1.99	(1.40)
N	3,300		3,121		2,336	

Group-level clustered standard errors in parenthesis. ***, **: significant at the 1% and 5% level

Table 11: Worker Marginal Effects From Fractional Regression Estimates

	Evade MFX		% of Compliance MFX	
	Coef	S.E	Coef	S.E
Free x Unknown Prob Unknown Fine - Follow Up	0.03	(0.09)	0.18	(0.14)
Free x Unknown Prob Known Fine	0.02	(0.07)	0.02	(0.12)
Free x Known Prob Unknown Fine	-0.15***	(0.04)	0.01	(0.16)
Restrict x Unknown Prob Unknown Fine - Follow Up	-0.06	(0.05)	-0.09	(0.10)
Restrict x Unknown Prob Known Fine	-0.12***	(0.04)	0.02	(0.13)
Restrict x Known Prob Unknown Fine	-0.10**	(0.05)	0.04	(0.16)
Public x Unknown Prob Unknown Fine - Follow Up	-0.06	(0.05)	0.12	(0.12)
Public x Unknown Prob Known Fine	0.04	(0.09)	0.03	(0.12)
Public x Known Prob Unknown Fine	-0.11***	(0.03)	-0.12	(0.10)
NoComm x Unknown Prob Unknown Fine - Follow Up	-0.04	(0.08)	0.07	(0.12)
NoComm x Unknown Prob Known Fine	0.02	(0.07)	0.13	(0.14)
NoComm x Known Prob Unknown Fine	-0.09	(0.05)	-0.05	(0.11)
NoComm x High Prob High Fine	-0.09**	(0.04)	0.17	(0.11)
NoComm x High Prob Low Fine	0.03	(0.08)	0.01	(0.11)
NoComm x Low Prob High Fine	0.17	(0.10)	0.08	(0.12)
Ability	-0.00	(0.00)	0.005**	(0.002)
Payment	-0.00005***	(0.00001)	-0.00007***	(0.00002)
Audited & Not Caught (t-1)	-0.07***	(0.01)	0.04	(0.03)
Audited & Caught (t-1)	0.49***	(0.05)	-0.08**	(0.04)
Early Audited & Not Caught	-0.21***	(0.02)	0.01	(0.03)
Early Audited & Caught	0.84***	(0.01)		
Worker x Lab	0.09***	(0.04)	0.06	(0.07)
Group Size	-0.02	(0.02)	-0.04	(0.03)
Extraversion	0.01	(0.01)	-0.00	(0.01)
Agreeableness	-0.01	(0.01)	0.01	(0.01)
Emotional Stability	0.01	(0.01)	0.01	(0.01)
Conscientiousness	0.00	(0.01)	0.02	(0.01)
Openness	0.00	(0.01)	-0.02	(0.01)
Risk Aversion	0.01	(0.01)	0.00	(0.01)
Personal Income	0.02	(0.02)	-0.03	(0.02)
Household Income	-0.01	(0.01)	0.00	(0.02)
Age	-0.00	(0.00)	-0.00	(0.00)
Male	0.01	(0.04)	-0.12**	(0.06)
Experience	-0.00	(0.00)	0.01	(0.01)
Predicted Y	0.18		0.43	
N	5,435		5,435	

Group-level clustered standard errors in parenthesis. ***, **: significant at the 1% and 5% level

Table 12: Student Marginal Effects From Fractional Regression Estimates

	Evade MFX		% of Compliance MFX	
	Coef	S.E	Coef	S.E
Free x Unknown Prob Unknown Fine - Follow Up	-0.21	(0.13)	-0.12	(0.12)
Free x Unknown Prob Known Fine	0.04	(0.15)	-0.00	(0.06)
Free x Known Prob Unknown Fine	0.14	(0.13)	-0.03	(0.09)
Restrict x Unknown Prob Unknown Fine - Follow Up	-0.21***	(0.06)	-0.18***	(0.04)
Restrict x Unknown Prob Known Fine	-0.01	(0.09)	0.17	(0.10)
Restrict x Known Prob Unknown Fine	-0.00	(0.08)	-0.15***	(0.05)
Public x Unknown Prob Unknown Fine - Follow Up	-0.04	(0.08)	0.02	(0.06)
Public x Unknown Prob Known Fine	-0.12	(0.09)	0.04	(0.08)
Public x Known Prob Unknown Fine	-0.12	(0.09)	-0.06	(0.08)
NoComm x Unknown Prob Unknown Fine - Follow Up	-0.13	(0.09)	0.05	(0.07)
NoComm x Unknown Prob Known Fine	-0.01	(0.08)	-0.01	(0.05)
NoComm x Known Prob Unknown Fine				
NoComm x High Prob High Fine	-0.24***	(0.08)	0.10	(0.09)
NoComm x High Prob Low Fine	-0.01	(0.09)	0.08	(0.12)
NoComm x Low Prob High Fine	-0.17**	(0.08)	0.09	(0.12)
Ability	-0.004**	(0.002)	0.006***	(0.001)
Payment	0.0002***	(0.00003)	-0.0003***	(0.00003)
Audited & Not Caught (t-1)	-0.12***	(0.03)	0.05	(0.03)
Audited & Caught (t-1)	0.39***	(0.04)	-0.14***	(0.03)
Early Audited & Not Caught	-0.54***	(0.02)	-0.00	(0.03)
Early Audited & Caught	0.63***	(0.02)		
Worker x Lab				
Group Size				
Extraversion	-0.00	(0.01)	0.01	(0.01)
Agreeableness	0.01	(0.01)	-0.00	(0.01)
Emotional Stability	0.03***	(0.01)	-0.01	(0.01)
Conscientiousness	0.00	(0.01)	-0.00	(0.01)
Openness	0.02	(0.01)	-0.01	(0.01)
Risk Aversion	0.01	(0.01)	-0.00	(0.01)
Personal Income				
Household Income				
Age	-0.02	(0.03)	-0.03	(0.03)
Male	-0.02	(0.04)	-0.13***	(0.04)
Experience				
Predicted Y	0.44		0.30	
N	5,502		5,502	

Group-level clustered standard errors in parenthesis. ***, **: significant at the 1% and 5% level

Appendix D: Independent Reviewer Comments

In this appendix we include the comments made by our panel of independent reviewers. The set of comments refers to an *earlier version* of the report, and the present version incorporates those comments. We would like to thank the reviewers for their comments on the report. Our reviewers were Professor Frank Cowell, Professor Hans-Theo Normann, Professor Martin Sefton, and Professor Joel Slemrod. To preserve anonymity of comments, we randomly assigned a letter to each reviewer.

Reviewer A

Comments on Final Report for HMRC Project on Tax Compliance Experiments
Below I comment on the experiments that were run, their results, and the interpretation of results. The most striking results concern differences between student and non-student samples; my comments are focused on these, and particularly on the conclusions that are drawn from these results.

The Experiments
An extensive set of experimental sessions have been conducted, with over a thousand participants in all. Roughly half the sessions use student participants in a lab environment. The other sessions use non-students, some in a lab environment and some using an online interface.

In all sessions groups of five participants take the role of taxpayers. Each taxpayer earns income from a real-effort task and then completes a tax return. The experimenter then conducts random audits and imposes fines on taxpayers found under-reporting income. This is repeated over a series of periods. Participants also complete a lottery choice task to elicit risk attitudes and complete a personality questionnaire.

The experiments manipulate the extent to which taxpayers can communicate with one another, whether or not taxpayers are informed of the audit probability, and whether or not taxpayers are informed of the fine, and whether a follow-up rule that targets non-compliers is used. A further set of treatments use the no-communication condition with known fine and audit rates and vary the audit rate and fine across sessions.

Results
Behaviour of students and non-students appears to be very different. For example, in the baseline case without communication student compliance is sensitive to incentives. This comes across clearly in Figure 3 and the accompanying statistical tests. In contrast, doubling the fine and audit rate has practically no effect on non-student tax compliance (Figure 3).

Some of the regression results suggest a more nuanced picture. For example, doubling the fine and audit rate does increase the rate of full compliance in both student and non-student samples. However, the general picture is that non-student tax compliance is relatively insensitive to treatments: in all treatments compliance exceeds 80% and only exceeds 90% in treatments where the fine is unknown (see Figure 2).

The difference between student and non-student responses contrasts with the recent study of Alm, Bloomquist and McKee (“On the External Validity of Tax Compliance

Appendix D: Independent Reviewer Comments

In this appendix we include the comments made by our panel of independent reviewers. The set of comments refers to an *earlier version* of the report, and the present version incorporates those comments. We would like to thank the reviewers for their comments on the report. Our reviewers were Professor Frank Cowell, Professor Hans-Theo Normann, Professor Martin Sefton, and Professor Joel Slemrod. To preserve anonymity of comments, we randomly assigned a letter to each reviewer.

Reviewer A

Comments on Final Report for HMRC Project on Tax Compliance Experiments
Below I comment on the experiments that were run, their results, and the interpretation of results. The most striking results concern differences between student and non-student samples; my comments are focused on these, and particularly on the conclusions that are drawn from these results.

The Experiments
An extensive set of experimental sessions have been conducted, with over a thousand participants in all. Roughly half the sessions use student participants in a lab environment. The other sessions use non-students, some in a lab environment and some using an online interface.

In all sessions groups of five participants take the role of taxpayers. Each taxpayer earns income from a real-effort task and then completes a tax return. The experimenter then conducts random audits and imposes fines on taxpayers found under-reporting income. This is repeated over a series of periods. Participants also complete a lottery choice task to elicit risk attitudes and complete a personality questionnaire.

The experiments manipulate the extent to which taxpayers can communicate with one another, whether or not taxpayers are informed of the audit probability, and whether or not taxpayers are informed of the fine, and whether a follow-up rule that targets non-compliers is used. A further set of treatments use the no-communication condition with known fine and audit rates and vary the audit rate and fine across sessions.

Results
Behaviour of students and non-students appears to be very different. For example, in the baseline case without communication student compliance is sensitive to incentives. This comes across clearly in Figure 3 and the accompanying statistical tests. In contrast, doubling the fine and audit rate has practically no effect on non-student tax compliance (Figure 3).

Some of the regression results suggest a more nuanced picture. For example, doubling the fine and audit rate does increase the rate of full compliance in both student and non-student samples. However, the general picture is that non-student tax compliance is relatively insensitive to treatments: in all treatments compliance exceeds 80% and only exceeds 90% in treatments where the fine is unknown (see Figure 2).

The difference between student and non-student responses contrasts with the recent study of Alm, Bloomquist and McKee (“On the External Validity of Tax Compliance

Experiments", Tulane University Working Paper, 2010). In another recent working paper ("Measuring, Explaining, and Controlling Tax Evasion: Lessons from Theory, Experiments, and Field Studies", Tulane University Working Paper) Alm states "A common criticism of experimental economics is that the student subjects typically used may not be representative of taxpayers. However, there is now much evidence that the experimental responses of students are seldom different than the responses of other subject pools." Clearly, the Exeter report paints a very different picture. I think the difference between samples is very important and worth careful consideration.

Interpretations of results
The explanation for differences favoured in the report is based on non-students having more experience in the tax payment process and having developed norms of compliance which they bring with them into the experiment, and which shape their experimental decisions (paragraph 6.15). Students on the other hand treat the experiment as a game from which they try to extract the maximum payoff (paragraph 6.14). The conclusion is that students do not form a good subject pool for tax compliance experiments (paragraph 6.12), and future experiments need to use non-student subjects if results are to be valid (paragraph 6.16). This may be the correct interpretation and conclusion, but I am not completely convinced of either.

First, regarding the reasons for differences between subject pools, there are many differences between the student and non-student sessions beyond the experiences of subjects with the tax payment process. One obvious difference is the lab versus online interface. That there are important differences between these is suggested by the issue of drop-outs: in the online experiments some participants decided to drop-out after reading instructions (paragraph4.19), whereas I suspect participants who attended a lab session stayed to the end. This could be because it is psychologically easier to drop out in a non-lab setting, or it could be because the non-lab setting offers more attractive alternative uses of time, but in any case, the setting is different. The fact that non-students also participated in lab sessions offers an opportunity to look into this more systematically. Do the data from lab experiments with non-students look more like the student data or online data? Were incentive effects observed in lab sessions with non-students? I don't think these questions can be answered adequately by simply including a lab versus online dummy variable in the regression analysis of non-student data. Instead the lab versus online sessions should be compared as separate treatments.

Other differences across student and non-student samples include: recruitment methods, show-up fees (perhaps the low show-up fee makes students more interested in pursuing additional earnings), the exchange rate between experimental currency units and pounds. Perhaps also the student sample have experience in participating in experiments and recognise more easily that the "tax authority" mentioned in the instructions is the experimenter, whereas there is more uncertainty about the status of the tax authority, and hence more cautious completion of tax returns, among the non-student sample. Thus there seem to be many alternatives to the socialization hypothesis. For evaluating the socialization hypothesis more information about the non-student sample would be useful: exactly how experienced are non-student participants in the tax payment process? How many pay tax on a PAYE basis, and how many file tax returns?

I also did not find completely convincing the conclusion that, because students and non-students deliver different data, student samples deliver invalid data and non-student samples deliver valid data for studying tax compliance. It may actually be true that a policy of increasing fines would increase compliance, in which case it isn't

so clear that the student data is invalid and non-student data valid. Also, I think it is important to recognise that the data are generated by subjects engaging with a particular task. If students approached this task as a game in which they should maximise their experimental payoff, doesn't this suggest that students saw through the tax framing of the task and saw the task for exactly what it was: an individual risky choice task? Perhaps if taxpayer behaviour reflects a pro-social dimension to paying taxes, and if the experimental task were different so that there was a pro-social dimension to paying taxes (for example if tax payments resulted in benefits to others, or to the group), students would engage with the task in a way that is more similar to the way taxpayers engage with tax compliance decisions. These are, of course, speculative comments. The key point however is that absent understanding of reasons for differences in behaviour across samples it is difficult to draw strong conclusions about the validity of data from each sample.

Reviewer B

Principal Claims
The important development in this study is the broadening of the experimental subject pool from the usual focus on students to the inclusion of workers. A particularly striking difference between the two subject groups is the workers' high compliance.

Fonseca and Myles make a second claim for their study. They state that the fact that income is earned is likely to reveal "actual" rather than strategic behaviour. However I am not fully convinced by this claim.

Important claims are made about the (lack of) effect of communication. Since the chat is not analysed (this is promised later) I am not persuaded by this finding.
I am not convinced by the "conjecture on norms" stated on page 4. Were the students younger than the workers? If so might this not have an important role to play in explaining the difference in behaviour?

Contexts
Full comparison with the Alm et al (2009, 2010) contributions in the US context would have been welcome.

A detailed reference back to the lessons learned from the Pilot study would have been useful. On a first reading of the final report it was not clear to me how, if at all, my comments on the pilot had been taken into account. Responses to my comments were subsequently provided and these are listed (in bold) alongside my main comments on the pilot (see page 3 below). In my view important lessons from the pilot were not fully learned for the main study.

Potential Biases.
In view of the high no-show rates I would have welcomed some analysis of those who decided not to show up or who did not continue (presumably something was known about them).

On page 17 the bias surely has to something to do with individual competence and ability to perform basic tasks?

Mistakes.
On page 21 it is noted that some respondents reported more than they earned (a phenomenon that happens in real life). It would have been good to have known more

about accuracy by participants and how much they understood of the details of the experiment to which they were supposed to have responded. There is a blurring of the distinction between what *is* known and what *could be* known. On page 13 “Known probability” may be a slight misnomer. The probability may have been knowable by all, but individuals might have made a mistake or misremembered; this is not tested.

Reviewer C

I find the experimental design well-crafted, and the results well-explained. The results are troubling for the external validity of experiments about tax compliance.

Your explanation for the differences in behavior are plausible but not verified by the experiments themselves. Indeed, the paper suggests several behavioral explanations of different observed phenomena that do not form a coherent, comprehensive story, leaving much future work to be done. I am intrigued by the explanation of a social custom of full compliance given that under the PAYE system most people do not file and do not have much opportunity to evade. I am not yet convinced by the conclusion that future tax compliance experiments should use non-student subjects--we learn that student and non-students respond differently but I'm not sure we learn that one has more external validity than the other.

Overall, a very nice piece of work that I learned a lot from.

Reviewer D

Summary
This report is about tax compliance of UK taxpayers. In addition to an analysis of the responsiveness to auditing rates and financial penalties, the goal of the study is the role of social networks may play when in disseminating information about audit instances and tax evasion cases.

To this end, the authors conduct an experiment both with actual UK tax payers (workers) as well as with students recruited from a typical laboratory subject pool (at the University of Exeter). The majority of subjects from the field did the experiment online whereas all student participants did the experiment at the FEELE laboratory of the University of Exeter. In order to control for this effect (online vs. laboratory experiment), some worker participants also did the experiment at the laboratory--- they travelled to the lab. The experiment essentially consists of (i) a real-effort task in which the participants generated their income; (ii) this income was subject to a tax which could be declared or not; (iii) possibly, participants were audited and had to pay a fine if they failed to fully declare their income.

The main results of the experiments are as follows: student participants evaded more frequently than the UK citizens in the field. If so, they also declared less income. Students also reacted more strongly to changes in the level of the fines levels than the non-student participants. Whereas students were unresponsive to a higher likelihood of being audited, the compliance rate of the field participants was unresponsive to both the fine and the audit rate.

Evaluation
I think experimental research in this area is particularly useful and can, as it does in the experiments reported here, lead to interesting insights. Experiments can uncover behavioural patterns which would be difficult to detect from audit reports in the field. The experiments reported here were professionally and meticulously designed and

conducted. The treatments covered are comprehensive and well thought out. I liked the fact that a real-effort task was used, and the specific task employed is appropriate. The elicitation of risk attitudes and personality data nicely enrich the data set. The results are intriguing and have policy implications.

Comments
One main result is that students report less income than workers. The finding is consistent with the majority of worker participants being *attitudinally compliant*. By contrast, many student participants perceive this experiment (as concluded in the report) as a money maximisation exercise. If so, (future) laboratory experiments with student participants are nevertheless useful because they can capture the behaviour of those tax payers who are not attitudinally compliant and instead (and perhaps rationally) underreport. So I think we can learn a lot from student subject pools in this and in future experiments.

The notion that the population consists of two types, attitudinally compliant tax payers who do not respond much to incentives and income maximising citizens who do, fits with the following result:

Page 21: "Awareness of the likelihood of being audited leads to non-complying students to declare a lower proportion of their income, but it has no effect on the rate of full tax compliance itself. It has no effect on compliance behaviour of workers."
This shows that the income maximizing part of the population may believe the auditing rate is higher than it actually is. So publicising the audit rate may turn out to be an adverse strategy for the HMRC. In addition, and even though no effect on attitudinally compliant workers was observed here, a crowding-out effect seems conceivable: publicising the audit rate may turn a compliant tax payer into an underreporting maximiser because the publication of the audit rate may cause the tax compliance to be perceived as a mere gamble where risk and risk attitudes matter.

6.7: "Interestingly, the effect of being audited and caught under-reporting in the results in a higher likelihood of evading later on." I think this is an interesting and important result, with a parallel from people who got caught free riding in public transport.

Minor Issues
I had difficulties understanding Figure 4: is this the distribution of compliance rates at the individual level?

The following point should be clarified. In the Summary at the beginning of the report, we find: *"Surprisingly, students were particularly responsive to increases in fine levels, and unresponsive to higher likelihood of being audited. The overall compliance rate of workers was unresponsive to either the fine rate or the audit rate. However, there was some evidence that workers were less likely to choose to evade when both the audit and fine rates were high.*" On page 21 (in the Results in grey), we learn: *"Workers respond to information about fines, while students respond to information about audit rates."* Perhaps these are two different issues (information about fines and audit rates vs. fine and audit levels) but this could be put more clearly.

Experiments", Tulane University Working Paper, 2010). In another recent working paper ("Measuring, Explaining, and Controlling Tax Evasion: Lessons from Theory, Experiments, and Field Studies", Tulane University Working Paper) Alm states "A common criticism of experimental economics is that the student subjects typically used may not be representative of taxpayers. However, there is now much evidence that the experimental responses of students are seldom different than the responses of other subject pools." Clearly, the Exeter report paints a very different picture. I think the difference between samples is very important and worth careful consideration.

Interpretations of results
The explanation for differences favoured in the report is based on non-students having more experience in the tax payment process and having developed norms of compliance which they bring with them into the experiment, and which shape their experimental decisions (paragraph 6.15). Students on the other hand treat the experiment as a game from which they try to extract the maximum payoff (paragraph 6.14). The conclusion is that students do not form a good subject pool for tax compliance experiments (paragraph 6.12), and future experiments need to use non-student subjects if results are to be valid (paragraph 6.16). This may be the correct interpretation and conclusion, but I am not completely convinced of either.

First, regarding the reasons for differences between subject pools, there are many differences between the student and non-student sessions beyond the experiences of subjects with the tax payment process. One obvious difference is the lab versus online interface. That there are important differences between these is suggested by the issue of drop-outs: in the online experiments some participants decided to drop-out after reading instructions (paragraph4.19), whereas I suspect participants who attended a lab session stayed to the end. This could be because it is psychologically easier to drop out in a non-lab setting, or it could be because the non-lab setting offers more attractive alternative uses of time, but in any case, the setting is different. The fact that non-students also participated in lab sessions offers an opportunity to look into this more systematically. Do the data from lab experiments with non-students look more like the student data or online data? Were incentive effects observed in lab sessions with non-students? I don't think these questions can be answered adequately by simply including a lab versus online dummy variable in the regression analysis of non-student data. Instead the lab versus online sessions should be compared as separate treatments.

Other differences across student and non-student samples include: recruitment methods, show-up fees (perhaps the low show-up fee makes students more interested in pursuing additional earnings), the exchange rate between experimental currency units and pounds. Perhaps also the student sample have experience in participating in experiments and recognise more easily that the "tax authority" mentioned in the instructions is the experimenter, whereas there is more uncertainty about the status of the tax authority, and hence more cautious completion of tax returns, among the non-student sample. Thus there seem to be many alternatives to the socialization hypothesis. For evaluating the socialization hypothesis more information about the non-student sample would be useful: exactly how experienced are non-student participants in the tax payment process? How many pay tax on a PAYE basis, and how many file tax returns?

I also did not find completely convincing the conclusion that, because students and non-students deliver different data, student samples deliver invalid data and non-student samples deliver valid data for studying tax compliance. It may actually be true that a policy of increasing fines would increase compliance, in which case it isn't

so clear that the student data is invalid and non-student data valid. Also, I think it is important to recognise that the data are generated by subjects engaging with a particular task. If students approached this task as a game in which they should maximise their experimental payoff, doesn't this suggest that students saw through the tax framing of the task and saw the task for exactly what it was: an individual risky choice task? Perhaps if taxpayer behaviour reflects a pro-social dimension to paying taxes, and if the experimental task were different so that there was a pro-social dimension to paying taxes (for example if tax payments resulted in benefits to others, or to the group), students would engage with the task in a way that is more similar to the way taxpayers engage with tax compliance decisions. These are, of course, speculative comments. The key point however is that absent understanding of reasons for differences in behaviour across samples it is difficult to draw strong conclusions about the validity of data from each sample.

Reviewer B

Principal Claims
The important development in this study is the broadening of the experimental subject pool from the usual focus on students to the inclusion of workers. A particularly striking difference between the two subject groups is the workers' high compliance.

Fonseca and Myles make a second claim for their study. They state that the fact that income is earned is likely to reveal "actual" rather than strategic behaviour. However I am not fully convinced by this claim.

Important claims are made about the (lack of) effect of communication. Since the chat is not analysed (this is promised later) I am not persuaded by this finding.
I am not convinced by the "conjecture on norms" stated on page 4. Were the students younger than the workers? If so might this not have an important role to play in explaining the difference in behaviour?

Contexts
Full comparison with the Alm et al (2009, 2010) contributions in the US context would have been welcome.

A detailed reference back to the lessons learned from the Pilot study would have been useful. On a first reading of the final report it was not clear to me how, if at all, my comments on the pilot had been taken into account. Responses to my comments were subsequently provided and these are listed (in bold) alongside my main comments on the pilot (see page 3 below). In my view important lessons from the pilot were not fully learned for the main study.

Potential Biases.
In view of the high no-show rates I would have welcomed some analysis of those who decided not to show up or who did not continue (presumably something was known about them).

On page 17 the bias surely has to something to do with individual competence and ability to perform basic tasks?

Mistakes.
On page 21 it is noted that some respondents reported more than they earned (a phenomenon that happens in real life). It would have been good to have known more

about accuracy by participants and how much they understood of the details of the experiment to which they were supposed to have responded. There is a blurring of the distinction between what *is* known and what *could be* known. On page 13 “Known probability” may be a slight misnomer. The probability may have been knowable by all, but individuals might have made a mistake or misremembered; this is not tested.

Reviewer C

I find the experimental design well-crafted, and the results well-explained. The results are troubling for the external validity of experiments about tax compliance.

Your explanation for the differences in behavior are plausible but not verified by the experiments themselves. Indeed, the paper suggests several behavioral explanations of different observed phenomena that do not form a coherent, comprehensive story, leaving much future work to be done. I am intrigued by the explanation of a social custom of full compliance given that under the PAYE system most people do not file and do not have much opportunity to evade. I am not yet convinced by the conclusion that future tax compliance experiments should use non-student subjects--we learn that student and non-students respond differently but I'm not sure we learn that one has more external validity than the other.

Overall, a very nice piece of work that I learned a lot from.

Reviewer D

Summary
This report is about tax compliance of UK taxpayers. In addition to an analysis of the responsiveness to auditing rates and financial penalties, the goal of the study is the role of social networks may play when in disseminating information about audit instances and tax evasion cases.

To this end, the authors conduct an experiment both with actual UK tax payers (workers) as well as with students recruited from a typical laboratory subject pool (at the University of Exeter). The majority of subjects from the field did the experiment online whereas all student participants did the experiment at the FEELE laboratory of the University of Exeter. In order to control for this effect (online vs. laboratory experiment), some worker participants also did the experiment at the laboratory--- they travelled to the lab. The experiment essentially consists of (i) a real-effort task in which the participants generated their income; (ii) this income was subject to a tax which could be declared or not; (iii) possibly, participants were audited and had to pay a fine if they failed to fully declare their income.

The main results of the experiments are as follows: student participants evaded more frequently than the UK citizens in the field. If so, they also declared less income. Students also reacted more strongly to changes in the level of the fines levels than the non-student participants. Whereas students were unresponsive to a higher likelihood of being audited, the compliance rate of the field participants was unresponsive to both the fine and the audit rate.

Evaluation
I think experimental research in this area is particularly useful and can, as it does in the experiments reported here, lead to interesting insights. Experiments can uncover behavioural patterns which would be difficult to detect from audit reports in the field. The experiments reported here were professionally and meticulously designed and

conducted. The treatments covered are comprehensive and well thought out. I liked the fact that a real-effort task was used, and the specific task employed is appropriate. The elicitation of risk attitudes and personality data nicely enrich the data set. The results are intriguing and have policy implications.

Comments
One main result is that students report less income than workers. The finding is consistent with the majority of worker participants being *attitudinally compliant*. By contrast, many student participants perceive this experiment (as concluded in the report) as a money maximisation exercise. If so, (future) laboratory experiments with student participants are nevertheless useful because they can capture the behaviour of those tax payers who are not attitudinally compliant and instead (and perhaps rationally) underreport. So I think we can learn a lot from student subject pools in this and in future experiments.

The notion that the population consists of two types, attitudinally compliant tax payers who do not respond much to incentives and income maximising citizens who do, fits with the following result:

Page 21: "Awareness of the likelihood of being audited leads to non-complying students to declare a lower proportion of their income, but it has no effect on the rate of full tax compliance itself. It has no effect on compliance behaviour of workers."
This shows that the income maximizing part of the population may believe the auditing rate is higher than it actually is. So publicising the audit rate may turn out to be an adverse strategy for the HMRC. In addition, and even though no effect on attitudinally compliant workers was observed here, a crowding-out effect seems conceivable: publicising the audit rate may turn a compliant tax payer into an underreporting maximiser because the publication of the audit rate may cause the tax compliance to be perceived as a mere gamble where risk and risk attitudes matter.

6.7: "Interestingly, the effect of being audited and caught under-reporting in the results in a higher likelihood of evading later on." I think this is an interesting and important result, with a parallel from people who got caught free riding in public transport.

Minor Issues
I had difficulties understanding Figure 4: is this the distribution of compliance rates at the individual level?

The following point should be clarified. In the Summary at the beginning of the report, we find: *"Surprisingly, students were particularly responsive to increases in fine levels, and unresponsive to higher likelihood of being audited. The overall compliance rate of workers was unresponsive to either the fine rate or the audit rate. However, there was some evidence that workers were less likely to choose to evade when both the audit and fine rates were high.*" On page 21 (in the Results in grey), we learn: *"Workers respond to information about fines, while students respond to information about audit rates."* Perhaps these are two different issues (information about fines and audit rates vs. fine and audit levels) but this could be put more clearly.