Do Students Behave Like Real Taxpayers? Experimental Evidence on Taxpayer Compliance from the Lab and From the Field*

Lawrence Choo

University of Exeter

Miguel Fonseca

University of Exeter

Gareth Myles

University of Exeter

Institute of Fiscal Studies

Do Students Behave Like Real Taxpayers? Experimental Evidence on Taxpayer Compliance From the Lab and From the Field*

Lawrence Choo†, Miguel A. Fonseca‡ and Gareth D. Myles§

April 30, 2014

Abstract

We report on data from a real-effort tax compliance experiment using three subject pools: students, who do not pay income tax; company employees, whose income is reported by a third party; and self-employed taxpayers, who are responsible for filing and payment. While compliance behaviour is unaffected by changes in the level of, or information about the audit probability, higher fines increase compliance. We find subject pool differences: self-assessed taxpayers are the most compliant, while students are the least compliant. Through a simple framing manipulation, we show that such differences are driven by norms of compliance from outside the lab.

Keywords: tax compliance, real effort, field experiment.

JEL classification numbers: C91, H26

†University of Exeter; Email: *cylc201@exeter.ac.uk*

‡Corresponding author. University of Exeter; Email: *m.a.fonseca@exeter.ac.uk*

§University of Exeter and Institute of Fiscal Studies; Email: *g.d.myles@exeter.ac.uk*

1

"In this world nothing can be said to be certain, except death and taxes." Benjamin Franklin, 1817

1   Introduction

Tax is the primary tool used by governments to finance public administration and public services. However, due to the high costs of monitoring compliance, tax evasion is as old a concept as tax itself. Tax evasion remains an economically important problem in modern economies: the tax gap, which is the non-received tax revenue in a fiscal year, is estimated to be \$450 billion in the United States in 2006 (IRS, 2012) and £35 billion in the United Kingdom in 2012 (HMRC, 2013).

The economic analysis of the tax compliance decision began with Allingham and Sandmo (1972) and Yitzhaki (1974). In this class of models, the taxpayer chooses the level of evasion which maximises her expected utility, and risk arises from the possibility that a random audit may be conducted by the tax authority. The Allingham-Sandmo-Yitzhaki model predicts that tax evasion will fall when either the penalty rate or the probability of being caught evading increase. However, when confronted with values of the audit probability and the penalty rates close to those observed in practice, the model predicts that all taxpayers should evade. This is contradicted by evidence of generally high levels of compliance in most western economies: despite the large size of the estimated tax gap in the US, it only amounts to about 17% of total tax liabilities.[1]

The discrepancy between the predictions of the model and the data led some to argue that high levels of compliance are due to psychological phenomena such as norms of compliance, tax morale, or patriotism. An alternative set of explanations is that in reality, taxpayers may not believe the audit probability is exogenous, or they may not know the actual audit probability — see Hashimzade, Myles, and Tran-Nam (2012) for a survey of the behavioural economics research applied to tax compliance. The latter case is relevant since in practice most taxpayers do not know the likelihood with which their tax return is audited by their country's tax authority. Most uneducated guesses are often an order of magnitude away from the actual audit rate. Relaxing the assumption of a known audit probability turns the model into a decision under ambiguity. In this

---

[1]There have been numerous extensions of this model, such as making labour supply endogenous, including a choice between employment in a formal and informal sectors, and increasing the complexity of the income tax (see the surveys of Pyle, 1991 and Sandmo, 2005) but the basic results are robust.

framework, taxpayers do not know the true probability of audit, but may have prior beliefs about what probability is most likely. If they are pessimistic, they will assign a high likelihood to a very high audit rate, which is consistent the high levels of compliance.

We report on a series of experiments testing the effect of norms of compliance on behavior by sampling our subject pool from three distinct populations: undergraduate students, who are the typical sample in economics experiments but have never paid income tax; individuals in full-time employment who pay income tax through a third-party reporting system; and individuals who are self-employed and therefore self-report their income tax liabilities to the tax authority. We manipulate two standard policy levers in the classic models of tax compliance: the audit probability and the fine for non-compliance. We also consider the case where the audit probability is unknown.

The experiment was implemented on a sample of 520 individuals, of whom 200 were students, 200 were individuals who pay tax through a third-party system, and 120 self-employed taxpayers who file a return annually. We found very large subject pool differences, both in the level of compliance, as well as the responsiveness to changes in experimental parameters. Students were the least compliant subject pool, but also the most responsive to treatment changes, particularly to ambiguity in the audit probability, as well as changes in the fine for non-compliance. Self-employed taxpayers and taxpayers who pay through third-party reporting were more compliant and mostly non-responsive to different conditions.

A post-experimental survey uncovered that the vast majority of self-employed individuals may have exhibited high compliance levels in the experiment due to norms of honesty and compliance, in the sense that the experimental framing led them to translate their real-world behaviour into the experimental task. To investigate the role of norms of tax compliance from outside the lab on behaviour, we conducted an additional treatment in which any reference to tax, audits and fines was removed from the experimental materials. Average compliance in this treatment was reduced by half in the self-assessed sample, as well as the other two samples, highlighting the importance of norms in determining compliance in the lab.

The remainder of the paper is organised as follows. Section 2 contextualises our work in the existing experimental literature on tax compliance experiments, both done in the lab and in the field. Section 3 outlines the theory and hypotheses underpinning the experiment, and section 4

describes the experiment. Section 5 presents the analysis and main results. Section 6 discusses the paper's results and Section 7 concludes the paper.

2 Previous Tax Compliance Experiments

Our study contributes to a longstanding literature on tax compliance experiments. The earliest experimental study of tax compliance was conducted by Friedland et al. (1978) and since that study a steady flow of contributions have followed. The typical experiment takes a group of university student subjects who must choose how much of a given income to declare to the tax authority. The experimenter can vary the probability of audit, the tax rate or the fine for non-compliance. These variables can be known to the subjects with certainty, or they can be uncertain. The basic experimental design has not changed a great deal in the 30 plus years since the literature started — see Alm and McKee (1998) and Fonseca and Myles (2012) for reviews. The literature finds a small, but positive elasticity of tax evasion with respect to audit rates, and a smaller and surprisingly also positive elasticity with respect to penalty rates.

The key advantage of laboratory experiments is that, unlike the field, the experimenter can accurately detect evasion, since income is perfectly observable in the lab. When conducting an empirical analysis on economics of crime, in whatever guise it may take, the econometrician is always impaired by the fact that she only works with data from those individuals who are caught. One never gets data on criminals who have never been caught, or those who cheated and then, for whatever reason, decided to stop. As such, we can never have measures of the deterrence aspect of fines, and only unreliable measures of the punitive effect.

There are two criticisms of laboratory experiments that, if taken at face value, limit the extent to which one can apply their findings to outside the lab. The first is the conceptual abstraction surrounding the task: there is typically little context surrounding the decisions subjects must make. The second is that the typical subject sample used in experiments may not be representative of the population. While some emphasise the role of financial incentives and argue that the validity of lab experiments in undiminished by the nature of the subject pool (Falk and Heckman, 2009), others claim that the putative control inherent to the lab may prove counterproductive if the task is inherently artificial to the subjects taking part in the experiment, and emphasise the importance

of experience with the environment of interest in determining the external validity of any findings (Harrison and List, 2004).

In the context of tax evasion, the latter criticism equates to asking: why should one study tax evasion using a set of individuals who have never paid income tax? It is surprising that the experimental literature on tax evasion has only recently started to address this issue. Gërxhani and Schram (2006) experimentally studied compliance in two different countries (the Netherlands and Albania), and they looked at five separate subject pools: high school students, university students, high school teachers, non-academic university personnel, and university lecturers. The amount of under-reported income was higher in the Netherlands than in Albania, and higher for pupils and students than for teachers. Increasing the audit probability did not affect evasion in Albania, but did reduce evasion in the Netherlands. Alm, Bloomquist and McKee (2013) compared the behaviour of undergraduate students to university staff and faculty, who pay their taxes through third-party reporting. They find students were less compliant than non-students, but had qualitatively similar responses to treatment effects. Bloomquist (2009) compares compliance behavior in the lab to behavior from random audits in the field and finds the two samples to be qualitatively similar.

In contrast to the experimental literature, there is a relative paucity of empirical work on tax compliance using field data. The emergence of randomised control trials (RCTs) in economics and their widespread use in policy has resulted in greater access to reliable data on tax evasion from the field. Slemrod et al. (2001) conducted an RCT on taxpayers in Minnesota. A letter was sent to a random subset of taxpayers who had filed a federal tax return during 1995, informing taxpayers that the return they would file that year would be "closely examined". The data on the tax returns on the individuals receiving the letter were made available for the year of the intervention and the preceding year. The results showed that the effect of the letter depended on the level of income: low and middle income taxpayers who received the letter increased their reported income relative to the control group. The increase in reported income was also dependent on the source of income (higher among taxpayers declaring trade and business income than those declaring farm income), which indicated the effect of opportunity to evade. The surprise result was that the reported tax liability of the high income treatment group fell sharply relative to the control group. The authors proposed that this could be explained by the incentive to reduce the probability of an audit when

the probability was less than one, as opposed to the belief that not all income would be discovered if audited for sure.

Kleven et al. (2011) report the results of an RCT in Denmark. The objective of the RCT was to ascertain the effectiveness of prior audits and different audit probabilities on reported income of individuals who pay their taxes either through a third-party reporting system or via self-reporting. The sample was 42,800 individuals in Denmark who were chosen to be representative of the population. In the initial year (2007) one half of the sample was randomly selected for rigorous audit treatment while the remainder were not audited. In the next year (2008) letters containing the threat of an audit was randomly sent to individuals in both groups. The individuals were not informed that they were part of an experiment. One group received a letter stating that an audit would certainly take place, a second group received a letter stating that half the group would be audited, and a third group received no letter. These different letters provided an exogenous variation in the probability of being audited. The effect of audits on future reported income was studied by comparing the audit and no-audit groups. This showed that audits had a strong positive impact on reported income in the following year. The effect of the probability of audit on reported income was analysed using the threat-of-audit letter and no-letter groups. They find that evasion rates are close to zero among those who use third-party reporting, and significantly higher among those who self-report. Prior audits, and higher probability of future audit has a positive effect on compliance on self-assessed taxpayers but not on individuals who pay through third-party reporting. Also, the effects were stronger for the threat of an audit for certain than for the threat that half the group would be audited.

The main shortcoming of the randomised controlled trial approach is that one cannot directly observe evasion, even if taxpayers are thoroughly audited. For example, cash transactions are, by their very nature, outside the scope of an audit, and unless a full audit of a company's account is done — which is beyond the usual *modus operandi* of most tax agencies — the full extent of evasion can never be measured. One relies on variations in reported income as a proxy of compliance: if on average, reported income goes up as a result of a policy intervention, that must be due to higher compliance, rather than any other external factor. However, we cannot infer the impact of the policy on the fraction of taxpayers in full compliance as well as the effect on the fraction of income

reported by those who do not fully comply.

Our paper complements both literatures by setting up a real effort experiment, where there is individual level variation in income as well as accumulated wealth, but where evasion can be accurately detected. We are therefore able to estimate how the propensity to evade reacts to changes in income as well as accumulated wealth, as per Slemrod et al. (2001). We are also able to study the role of social norms of compliance by examining the behaviour of different subject pools, in particular individuals who pay tax through third-party reporting and through self-assessment (like Kleven et al. 2011), as well as students who are the traditional subject pool in lab experiments.

3   Theory and Hypotheses

Following Yitzhaki (1974), the standard economic model of the compliance decision considers an individual taxpayer in a single-period setting. The taxpayer has a given amount of income, Y, which is not directly observed by the tax authority, and has to choose an amount, $X \leq Y$, of this income to declare. If the declaration is audited then the true level of income is revealed with certainty. The discovery of undeclared income, $Y - X$, results in the payment of tax on the undeclared income plus an additional fine at rate f on unpaid tax. After the declaration decision is made, one out of two potential states of the world is realised. In the state of the world in which there is not an audit, the taxpayer is left with disposable income Y^n, where

$$Y^n = Y - tX. \tag{1}$$

The level of disposable income in the state of the world in which there is an audit is equal to Y^c, which is defined as

$$Y^c = Y - tX - ft(Y - X). \tag{2}$$

3.1   Preferences

We model individual preferences using the model proposed by Chateauneuf, Eichberger and Grant (2007).[2] In this model, ambiguity causes individuals to be responsive to the best and worst possible

---

[2]This is a special case of Choquet Expected Utility preferences, whose axiomatic foundations were derived by Schmeidler (1989).

outcomes. Let $p \in \Omega$ be a state of nature, corresponding to the (possibly unknown) probability with which a taxpayer is audited. The decision-maker has a utility function defined as follows:

$$V(f) = \delta \left[(1 - \alpha)M_i + \alpha m_i \right] + (1 - \delta)E_\pi u_i(Y, X), \tag{3}$$

where $E_\pi u_i(Y, X)$ denotes the expected utility of decision-maker i with respect to the probability distribution π on Ω, $M_i = \max_{p \in \Omega} u_i(Y, X)$, and $m_i = \min_{p \in \Omega} u_i(Y, X)$. Consistent with the literature on tax compliance, we assume u_i is increasing and concave. In other words, the decision-maker maximises a convex combination of the expected utility, the highest utility and the lowest utility from a given act.

We can interpret π as the decision-maker's subjective belief about the true state of the nature. The effect of ambiguity manifests itself in the weight $\delta \in [0, 1]$ the decision-maker assigns to the best and worst outcomes. Note that if $\delta = 0$, the model reverts to subjective expected utility. The attitude to ambiguity is measured by the $\alpha \in [0, 1]$ parameter: an individual whose α parameter equals zero overweights the best possible outcome, while an individual whose α parameter equals one overweights the worst possible outcome.[3]

3.2 The Compliance Decision

The decision-maker will select X to maximise (3), where $m_i = Y^c$, $M_i = Y^n$ and $E_\pi u_i(Y, X) = p u_i(Y - tX - ft(Y - X)) + (1 - p)u_i(Y - tX)$.[4] Collecting terms and rearranging, this gives the following maximisation problem:

$$\max_{\{X\}} u_i(Y - tX - ft(Y - X)) \left[\delta\alpha + (1 - \delta)p \right] + u_i(Y - tX) \left[\delta(1 - \alpha) + (1 - \delta)(1 - p) \right] \tag{4}$$

To obtain a sufficient condition for there to be non-compliance, the marginal utility of income declaration must be negative when the decision-maker declares his income truthfully:

$$\left. \frac{\partial V(X)}{\partial X} \right|_{X=W} = u_i'(Y(1 - t)) \left[(ft - t)(\alpha\delta + (1 - \delta)p) - t(\delta(1 - \alpha) + (1 - \delta)(1 - p)) \right] < 0 \tag{5}$$

---

[3]An alternative interpretation of the α parameter is that it captures the extent to which the individual believes Nature (or in our case, the tax authority) is a benevolent or malevolent agent.

[4]Technically, $m_i = \lim_{p \to 1} E_\pi u_i(Y, X)$, and $M_i = \lim_{p \to 0} E_\pi u_i(Y, X)$

Collecting terms and rearranging gives:

$$f < \frac{1}{\delta(\alpha - p) + p} \tag{6}$$

Equation (6) allows us to state a number of predictions about the effect of changing the audit probability on behaviour. In the absence of ambiguity ($\delta = 0$), an increase in the audit probability always leads to lower levels of non-compliance, as per the standard Yitzakhi model, where the sufficient condition for evasion is independent of preferences. In the presence of ambiguity ($\delta > 0$), the effect of raising the probability of audit on behaviour is weakened, and it becomes heterogeneous, since δ is an individual-specific parameter. This leads to the first hypothesis of the paper.

Hypothesis 1: *Increasing a known probability of audit will lead to higher levels of compliance.*

Fixing the audit probability, the effect of changing the weight in ambiguity preferences will depend on how the decision-maker views ambiguity. If the weight the decision-maker puts on the worst possible outcome, α, is larger than his subjective belief about the probability of audit, p, then the decision-maker is *pessimistic*. The more sensitive a pessimistic decision-maker is to ambiguity (i.e. a higher δ), the higher the level of compliance for a given level of audit probability. Conversely, if α is smaller than p, then the decision-maker is *optimistic*, and increasing δ leads to lower compliance.[5] Existing survey and experimental evidence (Andreoni et al., 1998; Alm et al., 1992) suggests that subjects' beliefs about the audit probability are in excess of its actual value, which suggests individuals may be ambiguity averse in the context of a tax compliance decision. This constitutes the second hypothesis.

Hypothesis 2: *Making the probability of audit unknown will lead to higher levels of compliance.*

Our experimental design considers three distinct subject pools: students, taxpayers who pay taxes through third-party reporting, and self-assessed taxpayers. Individuals in each of the three subject pools have distinct experiences with the national tax authority, as well as norms of compliance, which likely emerge through professional social networks, as well as the history of audits.

---

[5]Snow and Warren (2006) derive qualitatively similar results using a model where decision-makers are uncertain about the true probability of audit, and hold beliefs over the probability of audit in the form of a probability distribution over Ω.

There are numerous ways in which social interaction can be introduced into the compliance decision. One way to do so is illustrated by the social custom model of Myles and Naylor (1996). This model assumes 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 the more out of step the non-compliant taxpayer is with the remainder of society.

In the context of our theoretical framework, Myles and Naylor (1996) introduce social custom through additional structure on u_i, which now takes the following form:

$$u_i(\cdot) = \begin{cases} u_i(Y, X) + bR(1 - \mu) + c & \text{if } X = Y \\ u_i(Y, X) & \text{if } X < Y \end{cases} \tag{7}$$

with $b \geq 0$ and $c \geq 0$. The additional utility from adhering to the social custom of honest tax payment is $bR(1 - \mu) + c$, where μ is the proportion of population evading tax and $R' > 0$. The parameters b and c represent the attitude of the individual taxpayer toward the social custom and can be expected to be different across taxpayers. This is a special case of the more general model of norms proposed by Krupka and Weber (2013).

Evasion will only take place if the expected utility from evasion is greater than that from honesty. The first point to note is that the model can have perfect compliance even when the expected financial gain 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 point to note is that the choice of whether to evade or not depends on the proportion of the population who are evading, μ. The choice is not made by the taxpayer in isolation but is the outcome of a process of social interaction.

The expected utility model predicts that there should be no difference in the behaviour of students and non-students in a compliance experiment. Behavioural models argue otherwise: 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 socialisation process through which taxpayers absorb the social custom. This opens the possibility that the different subject groups could have very different behaviours. In particular, Kleven et al. (2011) report lower compliance levels and higher responsiveness to changes in audit probabilities

among taxpayers who self-report their income than those who pay through third-party reporting. This is consistent with the Myles and Naylor (1996) model, to the extent that norms of compliance should be stronger among taxpayers who pay taxes through third-party reporting than among self-assessed taxpayers. This leads to the next experimental hypotheses:

Hypothesis 3: *Non-students will be more compliant and less responsive to changes in the probability of audit than students.*

Hypothesis 4: *Self-employed subjects will be less compliant and more responsive to changes in the probability of audit than subjects who pay their taxes through third-party reporting.*

In all models, increasing the fine for non-compliance unambiguously leads to higher compliance, which is our next hypothesis.

Hypothesis 5: *An increase in the fine for non-compliance will increase compliance.*

Our final hypothesis is based on prior empirical work, rather than theory. Kleven et al. (2011) report higher compliance among self-assessed taxpayers, following an audit. However, experimental data on repeated tax compliance tasks (Mittone, 2006; Kastlunger et al. 2009) finds the opposite result: average compliance drops in the experimental period immediately following an audit — the 'bomb-crater' effect. The lab data on the bomb-crater effect comes exclusively from student experiments, which suggests non-students may be less prone to this behavioural bias. We therefore formulate our next hypothesis based on the empirical evidence to date.

Hypothesis 6: *Average compliance following an audit will decrease among student subjects, remain constant with third-party reporting taxpayers and increase among self-assessed taxpayers.*

4   Experimental Design and Procedures

In this section we will outline the experimental design, as well as the procedures we employed to recruit the experimental subjects, as well as the protocol used for running the experiments for each of the three different samples. For ease of exposition, we will refer to the undergraduate student

sample as 'students', the third-party reporting taxpayer sample as 'PAYE', and the self-employed sample as 'self-assessed', to focus on their tax status, rather than their employment.[6]

4.1 The Experiment

The experiment involved three main parts: the first part consisted of the main experiment testing tax compliance. The second part was a risk aversion elicitation experiment (Holt and Laury, 2002) and the third part consisted of a number of questionnaires eliciting personality measures, as well as attitudes towards tax paying. The four stages of the experiment included income generation, declaration of income for tax purposes, auditing and payoff generation. A period in the experiment consisted of Stages 1-4 below, and there were 15 periods in total. The individual stages are:

Stage 1: Real-effort Task: Subjects performed a real-effort task for a fixed piece rate. This task was intended to induce a feeling of "ownership" of income.[7] The task consisted of a set of 48 sliders on a screen (Gill and Prowse, 2012.) Subjects earned a fixed payoff by placing each of the sliders at its halfway point.

Stage 2: Tax declaration: Subjects were told their Stage 1 income and had to declare their taxable income.

Stage 3: Auditing: Subjects were audited with a fixed probability p. If an audit occurred and subjects truthfully declared their taxable income, no penalty is levied. Otherwise, subjects paid the unpaid taxes in full plus a fine at rate f on unpaid tax.

Stage 4: Final Round Payoffs: Subjects received their final payoff for the round.

The three treatment variables are the probability of audit, the fine level, and the subject pool. We have four distinct probability conditions. Our baseline condition is P20, in which the probability of audit is equal to 20% and is public information. We study the effects of an increase in the audit probability using P40, in which the audit probability is equal to 40%. The UP treatment tackles the issue of ambiguity in the audit probability, by making it unknown to subjects, while being equal to 20% in practice. Finally, P20N is a payoff-equivalent version of P20, with a neutrally framed

---

[6]PAYE stands for Pay-As-You-Earn, the system used in the UK for withholding income tax payments by companies on their employees' behalf.

[7]Bühren and Kundt (2013) find evasion is significantly linked to cost of effort in a real effort tax compliance experiment.

	Audit Rate			
	P20N	P20	P40	UP
Fine level F100	30, 29, 27	35, 36, 31	35, 35, 30	30, 30, 32
F200		35, 35	35, 35	

Numbers are sample size for Student, PAYE and Self-Assessed subject pools.

Table 1: Experimental Design. .

set up. We also consider two separate fine levels: in F100, non-compliant subjects who are audited pay a fine of 100% of unpaid tax; in F200, that fine is equal to 200% of unpaid tax.

Table 1 outlines the experimental design, as well as the number of subjects from each sample that participated in each treatment. The first number in each cell in the number of student subjects, the second number refers to the number of PAYE subjects and the third number is the number of self-assessed subjects. We only collected data on the student and PAYE samples in the F200 conditions, as the compliance level among Self-Assessed was close to 100% in the F100 conditions.

4.2   The Student Sample

The student sample was recruited from a pool of voluntary undergraduate student subjects from a UK university through the ORSEE system (Greiner, 2004). All the sessions took place in the experimental laboratory of the university. Upon arrival at the laboratory, subjects were assigned to their seat; once everyone was seated, no communication was allowed between subjects. The experimenters informed subjects they could not answer any questions from this point onwards. The experiment was run using z-Tree (Fischbacher, 2007). Subjects were paid individually in cash at the end of the session. The average payment was £15.89, which included a show-up fee of £5.

4.3   The PAYE Sample

The majority of the PAYE sample was recruited from a pool of voluntary subjects run by a market research company, Saros Research. We also recruited PAYE taxpayers from businesses in the local area, as well as employees of the university. The subjects recruited by Saros Research are regularly paid to take part in focus group research and/or online surveys. To the best of our knowledge no

subjects had taken part in economics experiments prior to our study taking place. Saros Research triaged subjects through an initial questionnaire which asked a battery of questions including a question asking whether they were full-time residents in the UK for tax purposes, and another question asking for their tax status.

The subjects took part in the experiment from home or their place of work. To facilitate participation, we conducted sessions in the evenings between 6pm and 9pm. The experiment was run using z-Tree (Fischbacher, 2007). We provided subjects with software that connected their computer to our university servers. Subjects were asked to log on to the online system at a pre-designated time. PAYE subjects were paid through a bank transfer or through a cheque which was mailed to their home address.[8] The subset of PAYE subjects who resided in the university's area were recruited through ads and email. They travelled to the university laboratory to take part in the sessions, and they were paid in cash. The average payment was £36.03, which included a show-up fee of £20.

4.4 The Self-Assessed Sample

The self-assessed sample was recruited from a pool of voluntary subjects run by a market research company, ICM Research. Like the PAYE sample, these are regular paid subjects in market research who had never taken part in an economics experiment. The triage process was identical to that of the PAYE sample. Given the nature of the research and the subject pool, to minimise potential self-selection of subjects, as well as bias in choices in the experiment itself, ICM Research conducted all the recruitment and payment of subjects. Furthermore, we took extra measures to ensure anonymity of subjects, which were disclosed to subjects at the recruitment stage. Firstly, the researchers did not have any access to the names of subjects. Each subject was given a unique ID number, through which they would make their decisions. Only ICM Research could link names to ID numbers for payment purposes, but they could not access the experimental data itself.

To minimise direct contact with subjects, we designed a bespoke web-based software, which

---

[8]Upon signing up for the experiment, subjects provided us with their banking details through a secure web server, or with an address should they wish to be paid by cheque. Nobody declined to participate due to the method of payment.

had the same visual interface as the software used by student and PAYE subjects.[9] To access the experiment, subjects had to type their ID number plus a password. Subjects could log on at any time they wanted, within a week of receiving their log in information. However, once logged in to the experimental software, subjects had to complete the experiment within one hour of logging on. The experimental software did not allow subjects to log back in once an hour had elapsed. Subjects were paid by ICM via bank transfer. The average payment was £46.94, which included a show-up fee of £30.[10]

4.5 Experimental Procedures

Despite the differences in the recruitment of the three different subject pools, as well as the differences in the way they took part in the experiment itself (i.e. online vs. the lab), the actual protocol of the experiment was the same across the three samples. Upon logging on to the software, subjects had 10 minutes in which to read the instructions on their computer screen, after which the experiment started. Subjects could not interrupt the experiment and log back on at a later time. Each period had a fixed duration; after that time elapsed, the next period commenced until the end of the experiment. Once all three parts of the experiment were complete, a debrief text appeared on the screen, which explained the purpose of the experiment, and were given the option to opt out of the study if they wished to do so. Subjects were paid after finished reading the debrief form. The experiment lasted for no longer than one hour. All recruitment materials and instruction sets are available in the Appendix.

5 Results

The analysis will focus on the subjects' compliance rate, which we define as the ratio of declared income to income earned in a given period of the experiment. This definition means that the

---

[9]Note that the differences in the experimental design for this sample should lead, if anything, to *more* evasion among self-assessed subjects.

[10]The different show-up fees reflected the different opportunity cost of time for each sample. For that reason, we also implemented a different exchange rate between ECU and pound sterling depending on the sample: in the student sample 30 ECU equalled £1, whereas in the PAYE and Self-Assessed samples 15 ECU equalled £1.

compliance rate has a value between zero and one, which imposes some constraints on our method of data analysis when we go beyond analysing average treatment effects. We elaborate on this issue in the appropriate sub-section.[11] We begin the analysis of results by looking at the effect of the different treatments on the average compliance levels. We then proceed to econometrically estimate the determinants of compliance.

5.1 Average Compliance

Table 2 displays average compliance in the different treatments, for each of the three subject pools using the average behaviour of each subject over the course of the experiment as the unit of observation.[12] We start by examining the effect of increasing the audit probability, when it is known (i.e. treatments $P20$ and $P40$). With the exception of treatment F200 in the PAYE sample, where we observe a marginally significant difference ($t = 1.55, p = 0.063$), doubling the probability of audit has no statistically significant effect on average compliance, in all three subject pools. We now move to the effect of unknown audit rates on behaviour. When we compare behaviour in the treatment when audit rate is unknown (UP) to the treatments when the audit rate is known (P20, P40), we observe a marginally significant increase in average compliance levels in students (UP=P20: $t = 1.59, p = 0.058$), but no difference among PAYE and self-assessed subjects. This suggests that students are more sensitive to ambiguity than non-students.

Table 2 also reveals systematic differences in average compliance across the different subject

---

[11]About 4% of our data recorded subjects over-declaring their income. Unlike under-declarations, where it is impossible to distinguish between an individuals mistake and evasion, we can treat these observations as clearly errors and as such dropped those observations from the sample. While a frequent outcome is for a subject to make one mistake during the whole experiment, we found that 38% of over-declarations were made by 14 subjects (2.7% of the sample). We are confident that excluding these observations from the sample is simply ruling out the small subset of subjects who, despite our best efforts, perhaps did not understand the instructions quickly enough. Nevertheless our results would not qualitatively change if we had censored our dependent variable at 1.

[12]Given the large number of independent observations in our sample, we will employ the t test when testing for significant difference between average compliance levels in two treatments. This is because with sufficiently large samples, the distribution of the t-statistic asymptotically follows the Student's t distribution, even if normality is violated. We employ a conservative version of the two-sample t test which does not assume equality of variances in the two samples and allows for different sample sizes – see Sheskin (2011), pp 458-459 for a discussion. Since all our hypotheses are directional, unless otherwise noted, we will employ one-sided tests.

	Student				PAYE				Self-Assessed			
	P20N	P20	P40	UP	P20N	P20	P40	UP	P20N	P20	P40	UP
F100	0.30	0.61	0.64	0.73	0.51	0.84	0.80	0.84	0.44	0.93	0.93	0.93
	(0.23)	(0.35)	(0.36)	(0.30)	(0.34)	(0.24)	(0.23)	(0.23)	(0.30)	(0.19)	(0.15)	(0.16)
N	30	35	35	30	29	36	35	30	27	31	30	32
F200		0.79	0.84			0.81	0.89					
		(0.29)	(0.23)			(0.24)	(0.21)					
N		35	35			35	35					

Standard deviations in parentheses.

Table 2: Average Compliance Rate

pools. In the low fine treatments (F100), students exhibit significantly lower average compliance than both PAYE (P20, $t = 3.32, p < 0.001$; P40, $t = 2.17, p = 0.017$; UP, $t = 1.569, p = 0.061$) and Self-Assessed taxpayers (P20, $t = 4.77, p < 0.001$; P40, $t = 4.23, p < 0.001$; UP, $t = 3.19, p = 0.001$). There are less pronounced but significant differences in average compliance between PAYE and self-assessed subjects. Surprisingly, the latter subject pool is more compliant than the former (P20, $t = 1.72, p = 0.045$; P40, $t = 2.63, p = 0.005$; UP, $t = 1.74, p = 0.044$).

We conclude our discussion of Table 2 by looking at the effect of increasing fine levels. Doubling the level of fine led to significantly higher average compliance levels in the student sample in both audit probability conditions (P20, $t = 2.37, p = 0.010$; P40, $t = 2.79, p = 0.003$), as well as the PAYE sample, but only in the P40 condition (P20, $t = 0.53, p = 0.298$; RP40, $t = 1.78, p = 0.040$). This meant that the differences in average compliance in the two subjects pools in the low fine conditions disappear in the high fine conditions (F200-P20, $t = 0.36, p = 0.359$; F200-P40, $t = 0.93, p = 0.178$).

5.2 Distribution of Compliance

Restricting our analysis to treatment averages naturally ignores a great degree of heterogeneity in the data. Figure 1 shows the set of histograms of compliance levels, using the average compliance by each subject as the unit of observation. While our measure of compliance is a continuous variable, Figure 1 illustrates that when we aggregate individual compliance behaviour over the course of the
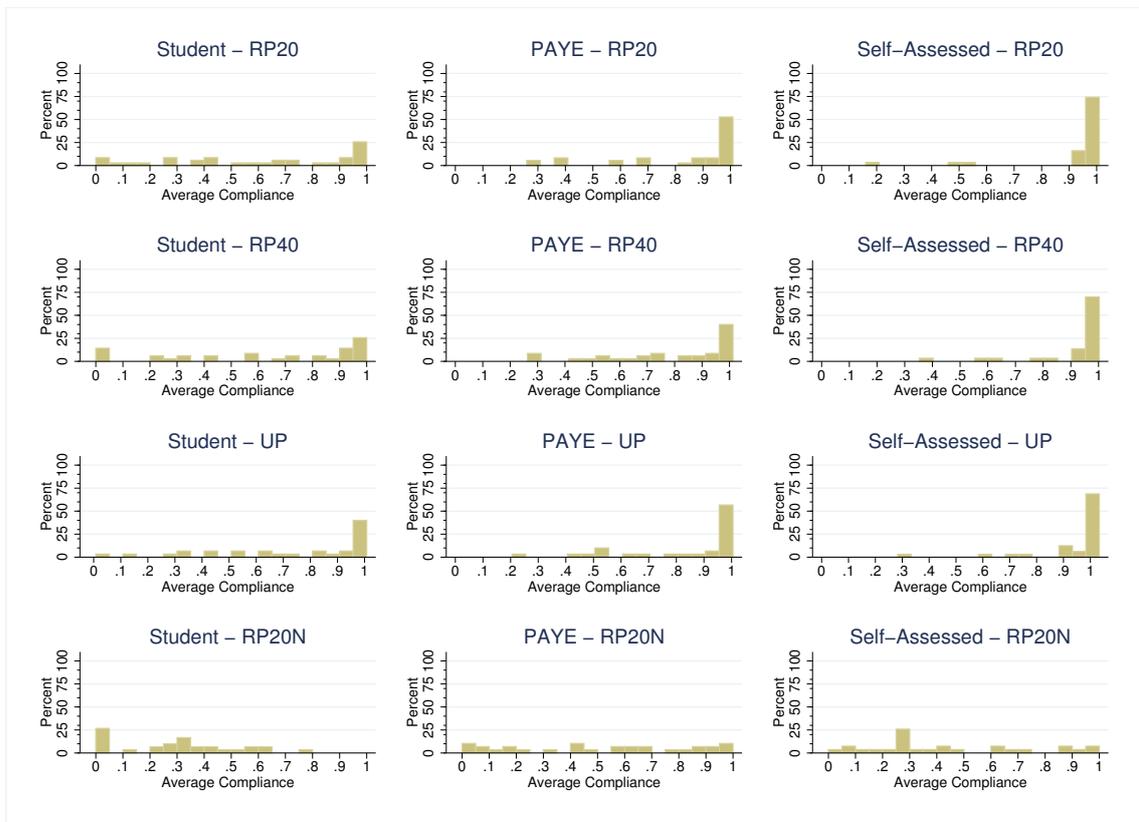
Figure 1: Histograms of Average Proportion of Declared Income, F100 Treatments

experiment, two behavioural types emerge: individuals who always declare their earnings in the experiment truthfully, and those who do not. Furthermore, the former category of behaviour is not only substantial, but it is also sensitive to treatment variations. The fraction of full compliers also substantially differs depending on the subject pool under consideration. In order to investigate this question we estimate a two-part model which explicitly distinguishes the two types of behaviour, and is therefore able to estimate the effect of our different treatments on the fraction of full compliers separately from the effect on behaviour of those who evade.

Let $c_i \in [0, 1]$ denote the average compliance level of individual i over the course of the experiment. We wish to estimate $E(c_i|x)$, where x is a vector of observables using a two-part model. We begin with the first part. Let $c_i^* = 0$ if $c_i = 1$, $c_i^* = 1$ if $c_i \in [0, 1)$. Then,

$$\Pr(c_i \in (0, 1)|x) = E(c_i^*|x) = F(x\beta_1) \tag{8}$$

where β_1 is a vector of variable coefficients and $F(.)$ is a cumulative distribution function to be specified.

The second part of our econometric model deals with subjects who, in some or all periods, reported a lower income than the one earned, and therefore have $c_i \in (0, 1)$.[13]

$$E(c_i|x, c_i^* = 1) = M(x\beta_2) \tag{9}$$

where $M(\cdot)$ is also a cumulative distribution function to be specified. As such, our two-part econometric model is defined as follows:

$$E(c_i|x) = E(c_i|x, c_i^* = 0) \cdot \Pr(c_i^* = 0|x) + E(c_i|x, c_i^* = 1) \cdot \Pr(c_i^* = 1|x)$$

As the first term of the left-hand side of the equation is zero, the model becomes:

$$E(c_i|x) = E(c_i|x, c_i^* = 1) \cdot \Pr(c_i^* = 1|x) = M(x\beta_2)F(x\beta_1) \tag{10}$$

From equation (10), we can derive the effect of a unit change in a regressor x_j on the conditional mean of c_i:

$$\frac{\partial E(c_i|x)}{\partial x_j} = \frac{\partial M(x\beta_2)}{\partial x_j}F(x\beta_1) + \frac{\partial F(x\beta_1)}{\partial x_j}M(x\beta_2) \tag{11}$$

---

[13]Subjects whose average compliance was zero account for less than 2% of the data. Estimating full non-compliance (i.e. $c_i = 0$) as a separate decision does not change the results and adds unnecessary complexity to the estimation.

The effect of a unit change in x_j will manifest itself on (i) the change in average compliance within the subset of subjects who do not fully declare their income, weighted by the proportion of those who evade; and on (ii) the change in the proportion of those who evade, weighted by the expected compliance level of those who evade. Since we are able to estimate $M(x\beta_2)$ and $F(x\beta_1)$ separately, we can report on partial effects of each part of the model. We estimate the binary part of the model, $F(x\beta_1)$, using the standard logit maximum likelihood estimator, and the fractional part of the model, $M(x\beta_2)$, using the logit quasi-maximum likelihood estimator for fractional models developed by Papke and Wooldrige (1996).[14]

Table 3 reports the average partial effects of the estimation of the model in its binary and its fractional part. The set of regressors consists of a set of dummies, each of which corresponds to an interaction between a treatment condition and a subject pool. The omitted category is the P20-Student treatment.[15] The coefficients on the variables in the binomial part of the model can be interpreted as the change in the likelihood of full compliance over the course of the experiment due to a change in the regressors (e.g. a different treatment condition and/or subject pool or individual characteristics. The coefficients on the variables in the fractional part of the model can be interpreted as the change in expected compliance resulting from a change in the regressors.

We do not observe a significant change in either the likelihood of full-compliance or in the average non-compliance in each of the three subject pools resulting from a change in the audit probability. This, together with the analysis of average compliance, forms our first result.

Result 1: *Doubling the audit probability results in no significant change in compliance in any of the three subject pools.*

We now turn to the effect of ambiguity in the audit probability. We find introducing ambiguity in the audit probability leads to no significant change in either the probability of full compliance

---

[14]See Ramalho et al. (2011) for a survey of the applications of fractional regression models. We could not reject the null hypothesis of misspecification for each part of the model using both the standard RESET test or the Goodness-of-Functional-Fit (GOFF) tests by Ramalho et al. (2013).

[15]We also estimated a different specification, where we included subject specific variables, such as number of years spent on current employment, risk aversion, age, a gender dummy, and a set of personality characteristics based on the Big-5 model. These individual-specific regressors were not significant, so we do not report them in the paper. Results from the estimations are available from the authors upon request.

	(1)		(2)	
DV	$c_i^* \in \{0,1\}$		$c_i \in (0,1)$	
P20 × PAYE	0.299***	(0.103)	0.143*	(0.076)
P20 × Self-Assessed	0.440***	(0.106)	0.254**	(0.119)
P40 × Student	0.036	(0.117)	0.023	(0.077)
P40 × PAYE	0.199*	(0.107)	0.143**	(0.069)
P40 × Self-Assessed	0.403**	(0.106)	0.280***	(0.094)
UP × Student	0.107	(0.116)	0.101	(0.078)
UP × PAYE	0.222**	(0.109)	0.203**	(0.081)
UP × Self-Assessed	0.422***	(0.105)	0.275***	(0.096)
P20 × F200 × Student	0.151	(0.109)	0.151*	(0.080)
P40 × F200 × Student	0.267**	(0.105)	0.177**	(0.079)
P20 × F200 × PAYE	0.151	(0.109)	0.186**	(0.074)
P40 × F200 × PAYE	0.310***	(0.104)	0.257***	(0.094)
P20N × Student	-2.664	(0.158)	-0.204***	(0.066)
P20N × PAYE	-0.111	(0.144)	-0.056	(0.073)
P20N × Self-Assessed	-0.181	(0.163)	-0.108	(0.069)
N	520		334	
AIC	1.168		1.062	

Robust standard errors in parenthesis.

***,**,*: significance at 1%, 5% and 10% level.

Table 3: Average partial effects of the determinants of average compliance

or the average non-compliance in all three subject pools.The attentive reader may wonder why our model does not pick up an effect, despite the noticeable change in the histograms for the the Student – UP and PAYE – UP conditions, namely an increase in the frequency of the top category. The reason for this slight increase is due to an increase in the fraction of subjects whose average contributions range between 0.96 and 0.99. However, this by itself is not sufficient to lead to a statistically significant effect, either in our econometric analysis in Table 3, or in the sample average analysis in Table 2; our ability to draw any conclusions about the effect of ambiguity on non-students in naturally restricted by the high compliance levels in the P20 and P40 treatments. This is our second result.

Result 2: *Making the audit probability ambiguous results in a small significant increase in students' average compliance relative to the treatment in which the audit probability was 20%, but not relative to the treatment where the audit probability was 40%. It resulted in no significant change in compliance for either PAYE or Self-Assessed taxpayers.*

We now compare behaviour across subject pools, keeping experimental parameters constant. We find that student subjects have a significantly lower likelihood of full compliance, as well as a lower expected compliance level by evaders relative to Self-Assessed taxpayers in all treatments.[16] When comparing students to PAYE taxpayers, with the exception of RP20-F100 ($\chi^2(1) = 7.93, p = 0.005$), we generally do not find a significant difference in the likelihood of full compliance. With regards to the expected compliance level by evaders, we find a significant difference when comparing behaviour in RP20-F100 ($\chi^2(1) = 3.52, p = 0.061$), but no significant difference in any other treatment.

Result 3: *Students exhibit lower average compliance than Self-Assessed taxpayers in all treatments and lower than PAYE taxpayers in the treatment with low fines and 20% audit probability. This difference is driven both by a lower likelihood of full compliance, and lower expected compliance levels by evaders.*

---

[16]Binary model: P20 $\times$ S-A = 0: $\chi^2(1) = 15.37, p < 0.001$; P40 $\times$ S-A = P40 $\times$ Stud: $\chi^2(1) = 15.58, p < 0.001$; UP $\times$ S-A = UP $\times$ Stud: $\chi^2(1) = 8.88, p = 0.003$. Fractional model: P20 $\times$ S-A = 0: $\chi^2(1) = 4.46, p = 0.035$; P40 $\times$ S-A = P40 $\times$ Stud: $\chi^2(1) = 6.83, p = 0.009$; UP $\times$ S-A = UP $\times$ Stud: $\chi^2(1) = 3.03, p = 0.082$.

We continue our comparison of behaviour across subject pools by focussing on the two non-student samples. We observe a statistically significantly higher likelihood of full compliance in the self-assessed sample than the PAYE sample in both RP40 ($\chi^2(1) = 4.32, p = 0.038$) and UP ($\chi^2(1) = 3.98, p = 0.046$), but not in RP20 ($\chi^2(1) = 2.12, p = 0.145$), but we find no difference in the expected compliance levels of evaders in either sample. This is our next result.

Result 4: *Self-assessed taxpayers exhibit higher average compliance than the PAYE sample. This difference is driven by a higher likelihood of full compliance.*

We complete our analysis by looking at the effect of doubling the fine rate. In the student sample, doubling the fine rate when the audit probability is low does not lead to a significant change in the likelihood of full compliance (P20 × F200 × Student =0: $\chi^2(1) = 1.89, p = 0.169$), but it does lead to significantly higher expected compliance among evaders (P20 × F200 × Student = 0: $\chi^2(1) = 3.55, p = 0.060$). When the audit probability is high, doubling the audit rate leads to significantly higher probability of full compliance (P20 × F200 × Student = P40 × F200 × Student: $\chi^2(1) = 5.01, p = 0.025$), as well as higher compliance among evaders (P20 × F200 × Student = P40 × F200 × Student: $\chi^2(1) = 3.44, p = 0.064$). In contrast, doubling the fine rate had no effect on behaviour in the PAYE sample.

Result 5: *Doubling the fine rate for non-compliance resulted in significantly higher average compliance among students, but only had a significant effect on behaviour by PAYE taxpayers when the audit probability was high.*

5.3 Dynamics of the Individual Compliance Decision

We now turn to the dynamic effects of auditing on compliance. To this effect, we will take advantage of the fact that the experiment was repeated multiple times on each subject, each of whom is a time series of compliance decisions. As such, we can employ standard methods of panel data econometrics to investigate the role of audits on future compliance.

Given that the compliance decision is bounded between zero and one and we are likely to observe mass points at either end of the distribution, standard GLS is not appropriate. As such, we opted

	(1)		(2)	
DV	$c_{it} \in [0,1]$		$c_{it} \in [0,1]$	
Constant	1.404***	(0.266)	1.340***	(0.312)
P20 × PAYE	0.727***	(0.222)	0.801***	(0.261)
P20 × Self-Assessed	1.120***	(0.250)	1.280***	(0.287)
P40 × Student	0.115	(0.188)	0.168	(0.227)
P40 × PAYE	0.439**	(0.214)	0.400	(0.257)
P40 × Self-Assessed	1.074***	(0.252)	1.178***	(0.291)
UP × Student	0.198	(0.367)	0.235	(0.443)
UP × PAYE	0.612***	(0.223)	0.663**	(0.264)
UP × Self-Assessed	1.096***	(0.242)	1.238***	(0.278)
P20 × F200 × Student	0.473**	(0.186)	0.570**	(0.224)
P40 × F200 × Student	0.714***	(0.193)	0.898***	(0.231)
P20 × F200 × PAYE	0.375*	(0.218)	0.348	(0.262)
P40 × F200 × PAYE	0.926***	(0.225)	1.009***	(0.262)
P20N × Student	-0.634***	(0.200)	-0.818***	(0.241)
P20N × PAYE	-0.276	(0.224)	-0.444*	(0.268)
P20N × Self-Assessed	-0.320	(0.254)	-0.501*	(0.302)
Income$_{it}$	0.010***	(0.002)	0.011***	(0.002)
Total Income$_{it-1}$	-0.001***	(0.0001)	-0.001***	(0.0001)
(Not Evade × Audited)$_{it-1}$	-0.337***	(0.031)		
(Evade × Audited)$_{it-1}$	-0.522***	(0.035)		
(Evade × Not Audited)$_{it-1}$	-0.226***	(0.035)		
Student × Audited$_{it-1}$			-0.386***	(0.032)
PAYE × Audited$_{it-1}$			-0.077**	(0.034)
Self-Assessed × Audited$_{it-1}$			-0.015	(0.050)
Experience$_i$	0.011*	(0.006)	0.013*	(0.007)
Risk$_i$	-0.016	(0.015)	-0.017	(0.018)
Male$_i$	-0.035	(0.091)	-0.030	(0.107)
Age$_i$	0.003	(0.005)	0.002	(0.006)
Extraversion$_i$	-0.014	(0.018)	-0.013	(0.021)
Agreeableness$_i$	-0.017	(0.018)	-0.021	(0.021)
Emotional Stability$_i$	-0.039**	(0.018)	-0.044**	(0.021)
Conscientiousness$_i$	-0.0001	(0.017)	-0.0003	(0.020)
Openness$_i$	-0.027	(0.019)	-0.032	(0.022)
N	5,900		5,900	
ρ	0.710	(0.023)	0.785	(0.017)

Standard errors in parenthesis. ***,**,*: significance at 1%, 5% and 10% level.

Table 4: Random-effects Tobit estimates of determinants of compliance

for a random effects two-limit Tobit model, in which:

$$c_{it} = x_{it}\beta + v_i + \varepsilon_{it} \qquad (12)$$

where x_{it} is a vector of regressors, β is the vector of coefficients to estimate, v_i are i.i.d. $N(0, \sigma_v^2)$ and ε_{it} are i.i.d. $N(0, \sigma_\varepsilon^2)$ independently of v_i. The observed data c_{it}^* is a potentially censored version of c_{it}. Our model will assume that $c_{it}^* = 0$ if $c_{it} < 0$, $c_{it}^* = c_{it}$ if $0 < c_{it} < 1$, and $c_{it}^* = 1$ if $c_{it} > 1$.

Table 4 summarises the estimates from the random effects Tobit estimations. We consider two separate specifications, which we explain below. The coefficient on the treatment dummies in both random effects Tobit estimations reiterate the findings from the analysis of average compliance: expected compliance is not sensitive to either changes or to ambiguity in the audit rate. Students are the least compliant subject pool and self-assessed taxpayers are the most compliant. Finally, doubling the fine rate leads to significant changes in compliance. Importantly, we can now investigate the dynamic aspects of the compliance decision, namely the effect of past audits on present compliance, as well as the effect of individual heterogeneity, whether manifested through different ability, accumulated wealth in the experiment, risk attitudes, or personality traits.

The individual-specific variables add very little explanatory power to the model; we can only marginally rejects the null hypothesis of no joint significance of all individual characteristic variables ($\chi^2(9) = 9.31, p = 0.098$). We find a significant coefficient on Experience, which measures the number of years in the current occupation (students had a value of zero), which has a small positive coefficient. The coefficient on Emotional Stability had a negative and significant coefficient. This is consistent with the evidence from Alaheto (2003), who found in a survey of convicted felons that emotional stability was negatively correlated with the likelihood of committing white collar crime.

The real-effort nature of the experimental design allows us to exploit individual differences in ability, which have a direct effect on the income each subject earned in a given period, as well as the accumulated income throughout the experiment. On one hand, we find a positive and significant coefficient on Income$_{it}$, which we interpret as evidence that higher ability subjects are less likely to evade.[17] On the other hand, the coefficient on Total Income$_{it-1}$ is negative and significant, which indicates a countervailing effect: the wealthier are our subjects, the more likely they are to evade.

---

[17]The reader will have noted from our description of the procedures, that the relative weight of the show-up fee on

We now focus on the effect of audits on behaviour. Regression (1) conditions the effect of an audit on whether a subject was fully complying or not, pooling across the three subject pools. To do so, we include three dummy variables interacting the decision of subject i in period t to evade or not (Evade$_{it}$; Not Evade$_{it}$) with the auditing outcome in period $t-1$ (Audited$_{it-1}$; Not Audited$_{it-1}$). The omitted category is the case where subject i did not evade in period t and was not audited in period $t-1$.

Starting with the case where the subject was not evading in period t, we observe a large negative and highly significant coefficient on Not Evade$_{it}$ × Audited$_{it-1}$, indicating that expected compliance goes down in the period subsequent to an audit taking place. The same occurs in the case where subjects are evading: not only are the coefficients on both Evade Not Audited$_{it-1}$ and Evade Audited$_{it-1}$ negative and significant, as expected, but there is a significant difference between the two coefficients, which indicates that even among evaders, the expected compliance goes down. Furthermore, the effect size of an audit on behaviour by non-compliant subjects (-0.185) is economically similar (though significantly smaller) to the effect of audits on compliant subjects.[18]

Having demonstrated that audits lead to lower compliance in the following auditing period, irrespective of whether one is a complier or evader, we wish to understand whether there are subject pool differences in the way compliance behaviour changes following an audit. Regression (2) tackles this problem by replacing the aforementioned audit interaction dummy variables with a new set of interactions between Audited$_{i,t-1}$ and a dummy for each subject pool. We find dramatic subject pool differences: while the coefficients on all interaction dummies are negative, we find a very large and significant coefficient on the student interaction dummy, and a smaller, though still significant coefficient on the PAYE interaction dummy. However, the Self-Assessed interaction is not significant. Furthermore, the coefficient on the student interaction is significantly larger than the coefficient on either of the non-student interaction dummies, but the latter two coefficients are not significantly different.[19] In short, we find evidence for the bomb-crater effect in our experiment,

total payment is different between students and non-students. This is primarily due to the fact that students were more effective at solving the slider task, and therefore had more income to declare. Our Income variable controls for that discrepancy.

[18]Not Evade Audited$_{it-1}$=0: $\chi^2(1) = 41.03, p < 0.001$; Evade Not Audited$_{i,t-1}$ = Evade Audited$_{it-1}$: $\chi^2(1) = 52.12, p < 0.001$; Not Evade Audited$_{i,t-1}$=Evade Not Audited$_{i,t-1}$-Evade Audited$_{it-1}$: $\chi^2(1) = 41.03, p < 0.001$.

[19]Student Audited$_{i,t-1}$ = PAYE Audited$_{i,t-1}$: $\chi^2(1) = 43.29, p < 0.001$; Student Audited$_{i,t-1}$ = Self-Assessed

but that effect is driven primarily by the student sample. We find weaker evidence of that effect among the PAYE taxpayer sample and no evidence among Self-Assessed taxpayers. This constitutes our next result.

Result 6: *Audits lead to a large fall in future compliance among students, as well as to a lesser extent, PAYE taxpayers. However, audits have no effect on future compliance behaviour of Self-Assessed taxpayers.*

5.4 Social Norms of Compliance

As mentioned before, we observed higher-than-expected compliance by the self-assessed taxpayers. To understand the reasons why that was the case, we followed up our experiment with a post experimental survey. Subjects were invited via ICM to fill out a 15-minute survey about a month after the data collection ended (see Appendix for the full set of questions).[20] It is impossible to correlate any survey responses to a specific individual in the experiment since the respondents in the survey were anonymous to the researchers. Of the 92 subjects who completed the experiments, 72 (85%) responded to the survey invitations.[21]

One potential reason why compliance levels were high among self-assessed taxpayers is because they were inexperienced subjects, and therefore declared their true earnings because they did not fully understand the instructions. The survey therefore started by inquiring about subjects' understanding of the rules of the experiment. About 80% of responders understood that they could declare a different level of income to that which they earned in a period, and roughly the same proportion stated they understood that they could potentially take more money home by under-declaring their income. As such, we can rule out the possibility that the large levels of compliance are due to misunderstanding of the rules of the experiment.

Audited$_{i,t-1}$: $\chi^2(1) = 38.98, p < 0.001$; PAYE Audited$_{i,t-1}$ = Self-Assessed Audited$_{i,t-1}$: $\chi^2(1) = 1.07, p = 0.302$.

[20]We collected the data on self-assessed sample several months after the Student and PAYE samples. As such, when we decided to collect a follow-up survey on the Self-Assessed sample, we could not replicate it on the other two samples.

[21]This does not include the subjects who took part in the UP treatment, and includes 31 subjects who took part in a related treatment, which did not fundamentally differ from the treatments presented in this paper. For details of that treatment, see Choo, Fonseca and Myles (2013).

Statement	Fraction of choices
I tried to always declare my income accurately.	0.81
I occasionally declared less income than I had earned.	0.09
I mostly declared less income than I had earned.	0.08
I always declared less income than I had earned.	0.01
I can't remember.	0.01

Table 5: Frequency of reported types of compliance

A second potential explanation for the high compliance rate is that subjects did not believe that the audit rate was the one stated in the instructions, or thought that it was not independent of their compliance behaviour. In the former case, 70% of respondents stated that the likelihood of being audited was the same as stated in the instructions. The large majority (63%) also stated that it was the same regardless of what they reported. However, a significant minority of subjects did not believe so. Some believed the likelihood of being audited increased if they reported a low income or under-declared their income, while others thought the audit likelihood was history-dependent. These responses may be a reflection of individuals' perceptions of the actual audit strategy taken by the tax authorities.

A third potential explanation is that norms of compliance drove the subjects' decisions. To understand the extent to which this was the case, we presented the following statements, summarised in Table 5, and asked subjects to indicate which statement best described how they behaved. Over 80% of respondents stated that they always declared their income accurately, and 18% of respondents stated declaring less than what they earned — 1% did not remember.

Subjects who stated always trying to accurately report their income were presented with a series of statements to which they could reply with 1 (Strongly agree) to 7 (Strongly disagree). In the following we present the percentage of subjects who had agreed (selected 1-3) with the selected statements, and we illustrate the data with comments made by those subjects when asked in an open-ended question to explain their approach to the experiment:

(90%): I declared all my income because it is the right thing to do.

"In real life I would be too concerned that I would be caught if I cheated on my tax return. I reflected this attitude in the experiment."

"It would not cross my mind to be intentionally dishonest."

"It's just the way I run my business. It's the easiest way"

(78%): I declared all my income because evasion is unfair on others.

"People moan about the state of the economy, but then do not declare all income. They have no right too, everybody pays we all have a better standard."

"I think I am an honest person, so only put down my earnings and I think everyone (individual and business) should pay their tax. If everyone paid the full amount we would all pay less. Too many big companies are riding on the backs of the UK public."

(73%): I declared all my income because that was the rules.

"It's just natural for me; even though I knew it wasn't 'real', I still found it very difficult to try to 'beat' the system."

Subjects who stated not always trying to accurately report their income were presented with a series of statements to which they could reply with 1 (Strongly agree) to 7 (Strongly disagree). In the following we present the percentage of subjects who had agreed (selected 1-3) with the selected statements, and we illustrate the data with comments made by those subjects when asked in an open-ended question to explain their approach to the experiment:

(71%) I took a calculated risk to not declare all my income.

"I thought it was the most profitable approach overall. Though I understood that I might incur penalties for understating the income earned, the scope for much greater justified (I think) the risk."

"I guess that it was weighing the probabilities in that after I had earned so much (c half way through) I reasoned that I could afford a few fines based on the income earned so far. Each time thereafter that i was not fined, I under-declared to the point that I was accurate up to halfway and not thereafter. It was a balance of probabilities call."

(78%) I wanted to earn as much from the experiment as possible.

29

"The potential fine was insufficient to counter the advantage of underdeclaring. Underdeclaring gave me a positive long run expectation."

(55%) I started to take more risks as the game wore on.

"Got away with it the 1st time so I continued to declare less amounts."

"Got a reasonable income from being truthful and then decided to see if I could boost it slightly by being dishonest."

"I was a bit slow off the mark. Once I worked out the actual penalties involved and the likely hood of being audited, I took the opinion that it was well worth the risk to under declare."

This evidence suggests that norms of compliance may have played a very important role in determining the compliance behaviour of the self-assessed taxpayers. Since we did not have matching survey data on Students and PAYE, in order to further understand the role of norms, we conducted an extra treatment with the same parameter values as F100-RP20, but in which the framing was neutral — we denote this treatment as P20N. That is, we removed all instances of tax, audit probability from the instructions and the text on the software interface, such that subjects were faced with the exact same decision problem, but without the normative context of compliance decision. We ran this treatment on the three subject pools.

Table 2 shows the average compliance in the neutral treatment was significantly lower in all three subject pools (students, $t = 4.12, p < 0.001$; PAYE, $t = 4.57, p < 0.001$; self-assessed, $t = 7.48, p < 0.001$). Furthermore, while students remained the least compliant subject pool on average in P20N, the average compliance level of self-assessed taxpayers is now lower than that of PAYE taxpayers (student vs PAYE: $t = 8.95, p < 0.001$; Student vs Self-Assessed: $t = 6.27, p < 0.001$; PAYE vs Self-Assessed: $t = 2.93, p = 0.002$). In our results from the estimation of the two-part model in Table 3, we find significant lower likelihood of full compliance, as well as a lower average compliance among evaders in the neutral treatment in all three subject pools.[22] Finally, in our

---

[22]Binary model, P20N × Student = 0: $\chi^2(1) = 836.66, p < 0.001$; P20N × PAYE = P20 × PAYE: $\chi^2(1) = 9.64, p = 0.002$; P20N × Self-Assessed = P20 × Self-Assessed: $\chi^2(1) = 15.50, p < 0.001$. Fractional model, P20N × Student = 0: $\chi^2(1) = 9.39, p = 0.002$; P20N × PAYE = P20 × PAYE: $\chi^2(1) = 6.82, p = 0.009$; P20N × Self-Assessed = P20 × Self-Assessed: $\chi^2(1) = 9.38, p = 0.002$.

analysis using the random effects Tobit estimations using period-level individual data reiterates our finding: compliance levels are lower in the P20N condition than in P20.[23]

In short, by removing the tax compliance context of the task, we turned a compliance decision into a choice under risk task, in a loss frame: a truthful declaration of earned income equals a sure loss, while under-declaration of income is a gamble in which detection yields a bigger loss than under full compliance and non-detection gives zero losses. We observe evidence in line with the literature on loss aversion, with a larger proportion of income allocated to the risky prospect than to the sure loss. We can therefore separate the role of risk preferences from the role of norms, and we demonstrate the importance of the latter in the decision whether or not to comply. This is our final result.

Result 7: *The high compliance level observed in our experiment, particularly in the Self-Assessed sample, can be attributed largely to norms of compliance.*

6 Discussion

The main finding of the paper is the stark behavioural differences between students and non-students: the former are less compliant but more responsive to policy levers than the latter. This difference seems to be primarily driven by norms of compliance, although our framing manipulation was not able to eliminate the differences in compliance between students and self-assessed taxpayers. It is possible that students are more loss averse than non-students, since we could not detect any role of risk aversion, at least insofar as our measure is able to do so. Self-assessed taxpayers and taxpayers who pay their taxes through third-party reporting seem to conform broadly to one of two types: compliant and non-compliant. The behaviour of the former seems to be primarily driven by norms of honesty, rather than financial incentives. This evidence is consistent with the models of norm compliance.

---

[23]Model (1), P20N $\times$ Student $= 0$: $\chi^2(1) = 10.11, p = 0.002$; P20N $\times$ PAYE $=$ P20 $\times$ PAYE: $\chi^2(1) = 21.50, p < 0.001$; P20N $\times$ Self-Assessed $=$ P20 $\times$ Self-Assessed: $\chi^2(1) = 39.02, p < 0.001$. Model (2), P20N $\times$ Student $= 0$: $\chi^2(1) = 11.26, p < 0.001$; P20N $\times$ PAYE $=$ P20 $\times$ PAYE: $\chi^2(1) = 23.40, p < 0.001$; P20N $\times$ Self-Assessed $=$ P20 $\times$ Self-Assessed: $\chi^2(1) = 45.90, p < 0.001$.

We also find significant subject pool differences in their behavioural response to an audit. While students' average compliance falls following an audit — the 'bomb-crater' effect — we register a qualitatively similar, but much smaller effect among third-party reporting taxpayers. However we observe no such changes in behaviour in the self-assessed subject pool. This finding is relevant when put into context of the lab and RCT literatures. On one hand, we demonstrate that this behavioural effect is broadly restricted to undergraduate students. On the other hand, we are able to reconcile the seemingly contradictory evidence from the two sets of randomised controlled trials on self-assessed taxpayers: Slemrod et al. (2001) report a sharp fall in reported tax liabilities after an audit among high-income groups after being exposed to the audit condition. In contrast, Kleven et al. (2011) report a *rise* in reported tax liabilities after an audit. The difference in these behaviours could be driven by taxpayers' beliefs about the actual audit rate, and whether that audit rate is exogenous. In our case where that audit rate is demonstrably exogenous, there is no subsequent change in compliance following an audit, since there is no reason why the compliance rate would change. This is unlike reality, in which taxpayers may have reason to believe the probability of being audited presently will strongly depend on their past behaviour.

Why do we observe such differences in compliance behaviour across subject pools? We start by discussing the role of procedural differences in the data collection. The primary difference is that the data from the student sample was collected in our laboratory, and both PAYE and self-assessed samples were collected online. While it is possible that this difference could lead to differences in behaviour, we argue that the expected effect is more compliance by subjects in the lab. There is a much larger social distance (the degree of reciprocity that subjects believe exist in a social interaction) between the subjects who took part in the experiment online and the experimenters vis-à-vis lab participants and experimenters.[24] Online subjects are not being directly observed while making their decisions, and they do not meet the experimenter face-to-face when they collect their payoff, which is directly impacted by the decision to evade. As such, they do not risk any hypothetical shame from facing the experimenter after having broken the norm of honesty. Furthermore, in the case of our self-assessed sample, we employed a double-blind design,

---

[24]Hoffman et al. (1996) study the impact of social distance on giving in dictator game and find higher social distance leads to more self-interested choices.

in which we did not know which subject took part in what treatment, and we used an intermediary to perform individual payments — something we made known to potential participants during the recruitment stage of the experiment. We also note that a subset our PAYE sample undertook the experiment in our laboratory, and their behaviour was also more compliant than that of students.

It is worthwhile comparing our data to audit data from the field. HMRC (2013) reported 37% of audited Business taxpayers and 14% of audited Non-Business taxpayers under-declared their tax liabilities (tables 6.4 and 6.5, pp. 38-39); the average reported income in our self-assessed sample was over 90%, but the percentage of self-assessed subjects who *always* correctly reported their income throughout the experiment was only equal to 65%. However, the tax gap associated with PAYE taxpayers in the UK is significantly smaller than that attributed to self-assessed taxpayers, while we find self-assessed participants no less compliant than PAYE in our experiment. We should note that comparisons between field data and our experiment should be made with caution for a number of reasons. Firstly, the filing decision in our experiment is an order of magnitude simpler than in reality, where self-assessed have multiple sources of income, as well as expenditures which they can use to offset their liabilities; furthermore, the type of income reported by self-assessed taxpayers is often different to that of PAYE taxpayers. Secondly, unlike our experiment, taxpayers in the field do not necessarily get randomly audited. Thirdly, the audit frequency in the field is an order of magnitude lower in the real world – we implemented a higher audit rate such the likelihood of being audited at least once during the experiment was reasonable. Finally, there is the issue of sampling: like the overwhelming majority of experiments, we relied on a pool of volunteer participants for the three samples, who may not be representative of the self-assessed taxpayer population.

It is also useful to compare our results to those of Alm et al. (2013), who study the behaviour of undergraduate students to that of university staff in the United States. That study also found that students were less compliant than staff, but they responded similarly to treatment changes. However, the mean compliance rates in their study are much lower than in our case. While it is possible to comment on the differences in levels of compliance in both studies, we cannot say much about treatment effects, since the two papers study different questions. Alm et al. (2013) report on treatments which measure the impact of information on the filing procedure when the process

itself is complex, as well as the role of tax deductions and tax credits. In that sense, the average compliance rate among students in baseline treatment in Alm et al. (2013), 0.618, is close to the average student compliance rate in our study, 0.61. We observe a larger difference in the compliance rate among non-students: 0.84-0.93 in our non-student samples compared to 0.795 in Alm et al. (2013). It is also worth pointing out that the subject pool differences between our study and Alm et al (2013) is not restricted to a difference in nationality: while most employees in the UK will not file a tax form at the end of the fiscal year, all taxpayers in the US must do so. In that sense, the sample of workers in Alm et al. (2013) is a hybrid of our PAYE and Self-Assessed samples. Replicating our design in the US, or the design by Alm et al. (2013) in the UK would be one way to verify whether cross-country differences are the reason behind this discrepancy. We leave this for future research.

7   Conclusion

Our experiment bridges two strands of the empirical literature on tax compliance: the longstanding literature on laboratory experiments and the nascent literature using randomised controlled trials. The former has relied primarily on undergraduate students as subjects and in most cases, taxable income was exogenously allocated, while in the latter the primary subject pool are actual taxpayers, whose taxable income is earned. We do so by conducting a laboratory experiment with three types of subjects who have very different experiences in the way they pay tax. We study the behaviour of students, who have no tax experience; workers who pay their income tax through third-party reporting systems; and self-employed individuals who file their own tax returns.

We find stark behavioural differences between students and non-students. The former are less compliant but more responsive to policy levers than the latter. Through a framing manipulation, we show that behavior in the three samples is driven by social norms of compliance, with the strongest effect coming from the Self-Assessed sample. A follow-up survey on the self-assessed sample corroborates the experimental findings. Our findings raise the obvious question of whether students are appropriate subjects for tax compliance experiments. Based on our evidence, under-graduate students may be more appropriate to study questions which require highly motivated, profit-maximising agents. Their behaviour will likely provide an upper bound for the effectiveness

34

of a particular policy tool that relies on financial penalties. Self-assessed taxpayers may be more appropriate to study questions pertaining to norms of compliance, and how such norms propagate through social and professional networks.

References

[1] Alaheto, T. 2003. "Economic crime: Does personality matter?" *International Journal of Offender Therapy and Comparative Criminology* 47, 335-355.

[2] Allingham, M.G., and Sandmo, A. 1972. "Income tax evasion: A theoretical analysis." *Journal of Public Economics* 1, 323-338.

[3] Alm, J., McClelland, G.H., and Schultz, W.D. 1992. "Why do people pay taxes?" *Journal of Public Economics* 48, 21-38.

[4] Alm, J., Bloomquist, K. M. and McKee, M. 2013. "On the external validity of laboratory policy experiments." *Mimeo*

[5] Andreoni, J., Erard, B., and Feinstein, J. 1998. "Tax compliance." *Journal of Economic Literature* 36, 818-860.

[6] Bloomquist, K. 2009. "A comparative analysis of reporting compliance behavior in laboratory experiments and random tax audits." National Tax Association Proceedings 2009, 113-122.

[7] Bühren, C., and Kundt, T.C. 2013. Worker or shirker – who evades more taxes? A real effort experiment. *MAGKS Joint Discussion Paper Series in Economics* No. 26-2013.

[8] Chateauneuf, A., Eichberger, J. and Grant, S. 2007. "Choice under uncertainty with the best and worst in mind: Neo-additive capacities." *Journal of Economic Theory* 137, 538-567.

[9] Falk, A., and Heckman, J. 2009. "Laboratory experiments are a major source of knowledge in the social sciences. *Science* 326, 535-538.

[10] Fischbacher, U. 2007. "Z-Tree - Zurich toolbox for Readymade Economic Experiments." *Experimental Economics* 10, 171-178.

[11] Friedland, N., Maital, S. and Rutenberg, A. 1978. "A simulation study of income taxation." *Journal of Public Economics* 10, 107-116.

[12] Gërxhani, K., and Schram, A. 2006. "Tax evasion and income source: a comparative experimental study." *Journal of Economic Psychology* 27, 402-422.

[13] Gill, D. and Prowse, V. 2012. "A structural analysis of disappointment aversion in a real effort competition." *American Economic Review* 102, 469-503.

[14] Greiner, B. 2004. "An online recruitment system for economic experiments." *MPRA Paper* 13513, University Library of Munich, Germany.

[15] Harrison, G.W., and List, J.A. 2004. "Field experiments." *Journal of Economic Literature* 42, 1009-1055.

[16] Hashimzade, N., Myles, G.D. and Tran-Nam, B. 2012 "Applications of behavioural economics to tax evasion." *Journal of Economic Surveys* 27(5), 941-977.

[17] HM Revenue and Customs. 2013. "Measuring tax gaps: 2013 edition." HMRC Corporate Communications.

[18] Hoffman, E., McCabe, K., and Smith, V. 1996. "Social distance and other-regarding behavior in dictator games." *American Economic Review* 86, 653-660.

[19] Holt, C.A., and Laury, S.K. 2002. "Risk aversion and incentive effects." *American Economic Review* 92, 1644-1655.

[20] Inland Revenue Service. 2012. "IRS releases new tax gap estimates; compliance rates remain statistically unchanged from previous study." available at `http://www.irs.gov/uac/IRS-Releases-New-Tax-Gap-Estimates;-Compliance-Rates-Remain-Statistically-Unchanged-From-Previous-Study`.

[21] Kastlunger, B., Kirchler, E., Mittone, L. and Pitters, J. 2009. "Sequences of audits, tax compliance, and taxpaying strategies." *Journal of Economic Psychology* 30, 405-418.

[22] Kleven, H.J., Knudsen, M.B., Kreiner, K.T., Pedersen, S., and Saez, E. 2011. "Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark." *Econometrica* 79, 651-692.

[23] Krupka, E.L., and Weber, R. 2013. "Identifying social norms using coordination games: Why does dictator game sharing vary?" *Journal of the European Economic Association* 11, 495-524.

[24] Mittone, L. 2006. "Dynamic behaviour in tax evasion: An experimental approach." *The Journal of Socio-Economics* 35, 813-835.

[25] 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.

[26] Papke, L. and Wooldridge, J.M. 1996. "Econometric methods for fractional response variables with an application to 401(K) plan participation rates." *Journal of Applied Econometrics* 11, 619-32.

[27] Pyle, D.J. 1991. "The economics of taxpayer compliance." *Journal of Economic Surveys* 5, 163-198.

[28] Ramalho, E.A., Ramalho, J.J.S., and Murteira, J.M.R. 2011. "Alternative estimating and testing empirical strategies for fractional regression models." *Journal of Economic Surveys* 25, 19-68.

[29] Ramalho, E.A., Ramalho, J.J.S., and Murteira, J.M.R. 2013. "A generalized goodness-of-functional form test for binary and fractional regression models." *The Manchester School.* Forthcoming.

[30] Sandmo, A. 2005 "The theory of tax evasion: A retrospective view." *National Tax Journal* 58, 643-663.

[31] Schmeidler, D. 1989. "Subjective probability and expected utility without additivity." *Econometrica* 57, 571-87.

[32] Sheskin, D.J. 2011. *Handbook of Parametric and Nonparametric Statistical Procedures.* London: CRC Press.

[33] Slemrod, J., Blumenthal, M., and Christian, C. 2001. "Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota." *Journal of Public Economics* 79, 455-483.

[34] Snow, A. and Warren Jr, R.S. 2006. "Ambiguity about audit probability, tax compliance and taxpayer welfare." *Economic Inquiry* 43, 865-871.

[35] Yitzhaki, S. 1974. "A note on income tax evasion: a theoretical analysis." *Journal of Public Economics* 3, 201-202.

A Recruitment and Experimental Materials

A.1 Recruitment Script

Dear panellist,

ICM Research is currently recruiting people to take part in a research study which will be conducted by the University of Exeter. The project is a different method of market research to that which you are used to and involves taking part in an online exercise about financial and economic decision-making.

You would be required to take part in an online activity between [DATE] and [DATE] which will take no more than 60 minutes to complete. You will be required to log in to a website in order to take part in the session. You can access the online activity from wherever is easiest for you, and at any time on the specified time frame. However, once you start, you must complete the exercise in one sitting. If you agree to participate ICM will send you a website address, username and password to access the research activity which is being hosted by the University of Exeter.

If you agree to take part, you will receive a cheque for 30 following completion of the online activity to say thank you for your time. Theres also the possibility of earning more throughout the activity itself. The online activity will consist of interactive decision-making. You will be faced with a particular scenario, and you will be asked to make a series of decisions.

Each decision has a direct bearing on the payment you will receive for participating. As such, depending on what you decide during the 60 minutes, your earnings could rise up to 60 from your

guaranteed 30. You will be told what the decisions will be before the activity starts.

ICM would like to assure you that all the information we collect will be kept in the strictest confidence. The University of Exeter will not be given your name or address and your decisions you make during the activity will be strictly anonymous. It will not be possible to identify any particular individual or address in the results.

We hope that you will agree to take part.

Kind regards,

ICM Research, University of Exeter

- Yes, I would like to take part in the research.

- No, I would not like to take part in the research

A.2 PROFILE QUESTIONS

Thank you for agreeing to take part. We just need to check a few details with you first.

- Gender

- Age

- Government Office Region

Q. Are you eligible to pay Income tax in the UK?

- Yes*

- No

- Dont know

Q. Which of the following best describes your work status? SINGLE CODE

- Employee full-time (30> hours per week)

- Employee part-time (<30 hours per week)

- Self-employed, full or part-time*

- On a government supported training scheme (e.g. Modern Apprenticeship/ Training for Work)

- Full-time education at school, college or university

- Unemployed and available for work

- Permanently sick/disabled

- Wholly retired from work

- Looking after the home

- Doing something else

*Note: Subjects who answered options marked with an * were subsequently invited to participate in the experiment.*

A.3 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 some of the factors that effect taxpayers decisions on paying tax. This was accomplished by asking participants to perform the slider task, which determined your pre-tax earnings. We then measured tax compliance by calculating the fraction of your pre-tax earnings you declared.

The purpose of this study was to see what effect changing audit rates has on tax compliance. To do this we varied the rules of the experiment from session to session: some had consistent audit rates in others the rates changes, in some sessions audits were random in others they depended on what earnings were declared. By doing this, we hope to gain a better understanding of taxpayer behaviour.

This research is funded by HMRC, the government department responsible for collecting and administering taxes in the UK. 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. HMRC often commissions independent organisations like Exeter University to undertake studies like this one. This study is part of a broader programme of research that HMRC undertakes to better understand the needs of its customers, its operational performance and how to make improvements to its service.

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. As you know, you were recruited by ICM research, who referred you on to a website designed by Exeter University. The study has been designed in such a way that Exeter University or HMRC cannot know who has taken part, and ICM research cannot know how a particular participants behaved during the study.

HMRC does not have access to any of your personal data associated with this study; they will only be given the anonymous data from the experiment, which will not have any names or other identifying information. It is not possible for HMRC to know who has taken part in the study at all.

We will publish 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. We will never report the results of individual participants. You can see the results from other research commissioned by HMRC on their website: http://www.hmrc.gov.uk/research/reports.htm

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:

A.4 Instruction Sets - Framed

Note: to economise on space, we will present only one version of the framed instruction sets. The instructions were identical across treatments, except for the description of the audit rate. We present the three versions of the sentence describing the audit rate, which we emphasise from the rest of the document by underlining them. We remind the reader that subjects only saw one version.

Welcome to our experiment. You will have 10 minutes to read these instructions; please read them carefully because you will NOT be able to refer back to them once the 10 minutes are up. Your cash earnings in this experiment will depend on the 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 earnings will be denoted in Experimental Currency Units (ECU). 15 ECU are worth £1. After the experiment is over, we will convert the total of your earnings for the whole session into pounds and pay you through a bank transfer. In addition to your earnings during the session, you will receive £30 for participating.

Please click Next to see the next page of instructions.
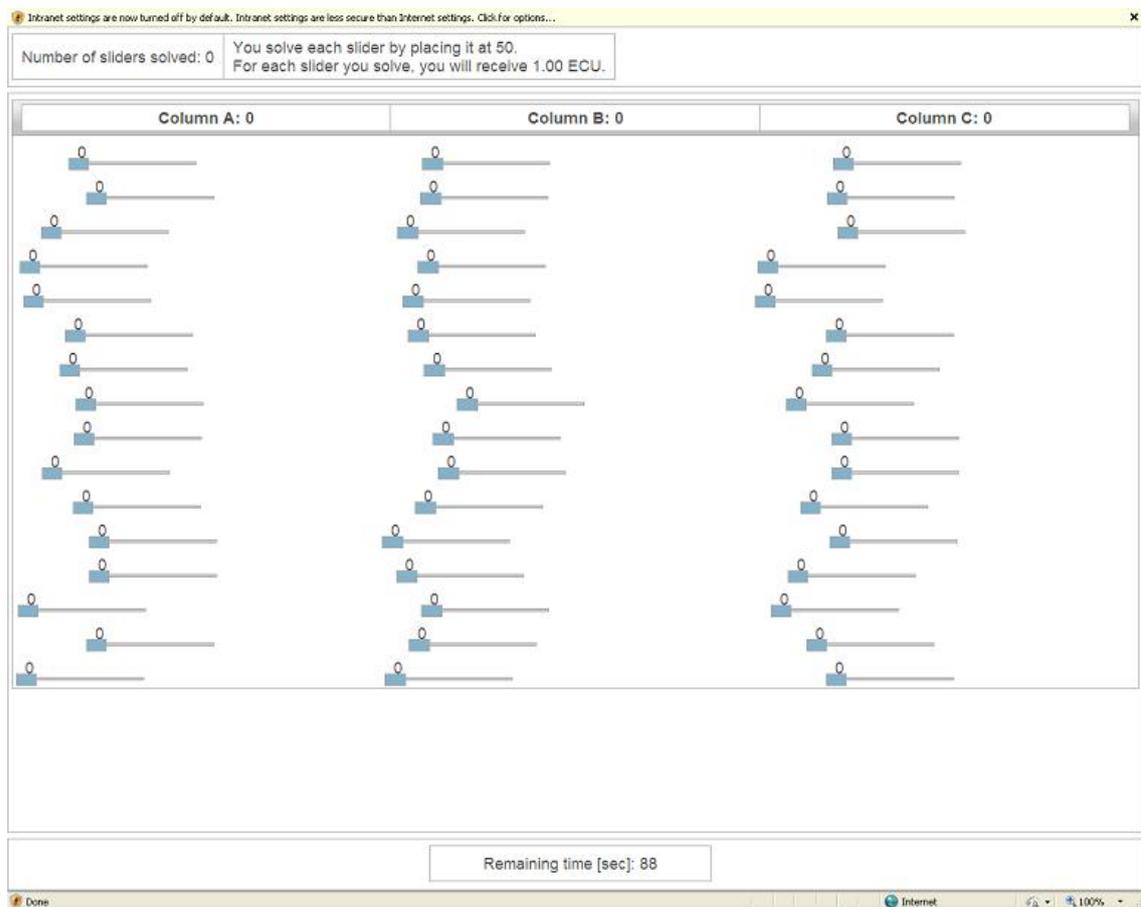
PART A

This part of the experiment is divided into 15 rounds. In each round, you will have the opportunity to earn ECU by solving as many sliders as possible within 100 seconds.

You will be presented with a screen with 48 sliders, and your task in each experimental round is simply to solve as many sliders as you can within the time limit.

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 on the top, 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. You will earn 1 ECU for each slider you solve.

Please click Next to see a screen shot of the slider task.

After 100 seconds, the task ends and the computer will inform you of the total number of sliders you had successfully solved as well as your total earnings from the task. This is your income from your task.

Your income is taxable and you will be required to fill in 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.

To fill in the tax return, you have simply to declare your earnings. Given the amount of earnings that you declare, the computer will automatically compute the total amount of tax you will be required to pay.

Please click Next to see a picture of the tax return you must fill in.

Once you complete your tax form, the tax authority may choose to audit it.

RP20 treatment: There is a 1 in 5 chance (or 20% probability) your tax form will be audited.
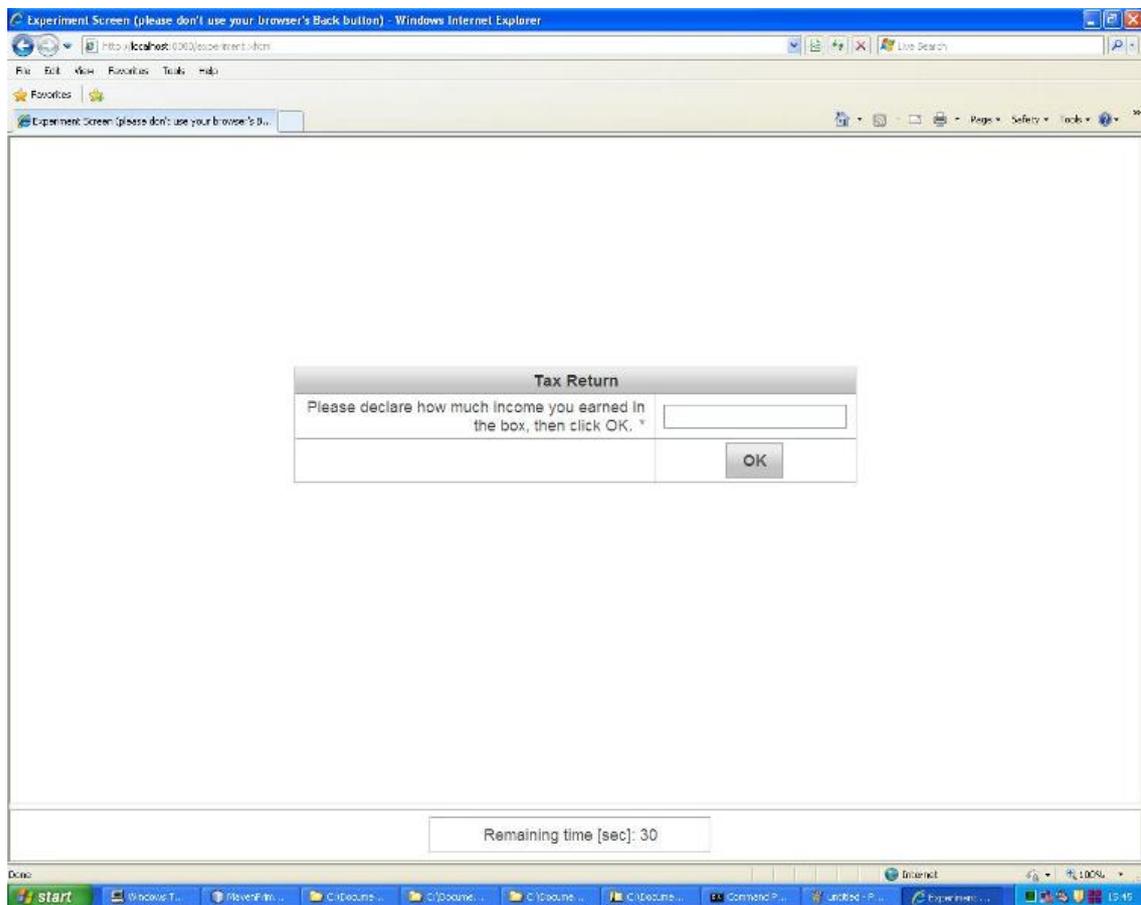
RP40 treatment: There is a 2 in 5 chance (or 40% probability) your tax form will be audited.

UP treatment: The chance your tax form will be audited is unknown.

If you are NOT selected for tax audit:
Your final earnings for the experimental round will be equal to the amount of earnings you had made from the task minus the tax you had paid on the earnings you reported on your tax form.
If you ARE selected for tax audit:

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

- F100 treatments If you are audited and if you reported less income than you actually earned, then you will be required to pay the extra amount of tax due to the authority. In addition

44

you will pay a fine of 1 ECU for each ECU of underpaid tax.

- <u>F200 treatments</u> If you are audited and if you reported less income than you actually earned, then you will be required to pay the extra amount of tax due to the authority. In addition you will pay a fine of 2 ECU for each ECU of underpaid tax.

- 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.

Now lets go through a few hypothetical examples.

Example 1:

Your Details	ECU
Your Income:	10.00 ECU
Your Declared Income:	10.00 ECU
(-) Tax Paid on Declared Income:	2.50 ECU
Selected by Tax Authorities for Audit?	Yes
(-) Additional Tax Payable (if any)	0 ECU
(-) Fine (if any)	0 ECU
(+) Tax refunded (if any)	0 ECU
END OF ROUND INCOME	**7.50 ECU**

Example 2:

Your Details	ECU
Your Income:	20.00 ECU
Your Declared Income:	16.00 ECU
(-) Tax Paid on Declared Income:	4.00 ECU
Selected by Tax Authorities for Audit?	Yes
(-) Additional Tax Payable (if any)	1.00 ECU
(-) Fine (if any)	1.00 ECU
(+) Tax refunded (if any)	0 ECU
END OF ROUND INCOME	14.00 ECU

Example 3:

Your Details	ECU
Your Income:	16.00 ECU
Your Declared Income:	10.00 ECU
(-) Tax Paid on Declared Income:	2.50 ECU
Selected by Tax Authorities for Audit?	No
(-) Additional Tax Payable (if any)	0.00 ECU
(-) Fine (if any)	0.00 ECU
(+) Tax refunded (if any)	0 ECU
END OF ROUND INCOME	13.50 ECU

After the audit is completed, you will be presented with a screen that summarises the following:

- 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 income 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 round.

Summary

In short, each round in this part of the experiment will consist of three stages:

STAGE 1: You earn income by solving sliders.

STAGE 2: You declare your income to the tax authority.

STAGE 3: You get your final earnings, which depend on how your report your income and whether you were audited or not.

This completes the description of Part A. The experiment will start automatically after 10 minutes, when the Remaining time [sec] hits zero.

A.5 Instruction Sets - Neutral

Welcome to our experiment. You will have 10 minutes to read these instructions; please read them carefully because you will NOT be able to refer back to them once the 10 minutes are up. Your cash earnings in this experiment will depend on the 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 earnings will be denoted in Experimental Currency Units (ECU). 15 ECU are worth £1. After the experiment is over, we will convert the total of your earnings for the whole session into pounds and pay you through a bank transfer. In addition to your earnings during the session, you will receive £30 for participating.

Please click Next to see the next page of instructions.

PART A

This part of the experiment is divided into 15 rounds. In each round, you will have the opportunity to earn ECU by solving as many sliders as possible within 100 seconds.

You will be presented with a screen with 48 sliders, and your task in each experimental round is simply to solve as many sliders as you can within the time limit.

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 on the top, 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. You will earn 1 ECU for each slider you solve.

Please click Next to see a screen shot of the slider task.

After 100 seconds, the task ends and the computer will inform you of the total number of sliders you had successfully solved as well as your total earnings from the task. This is your income from your task.

Upon learning how much income you have made, you must decide how to allocate your income.
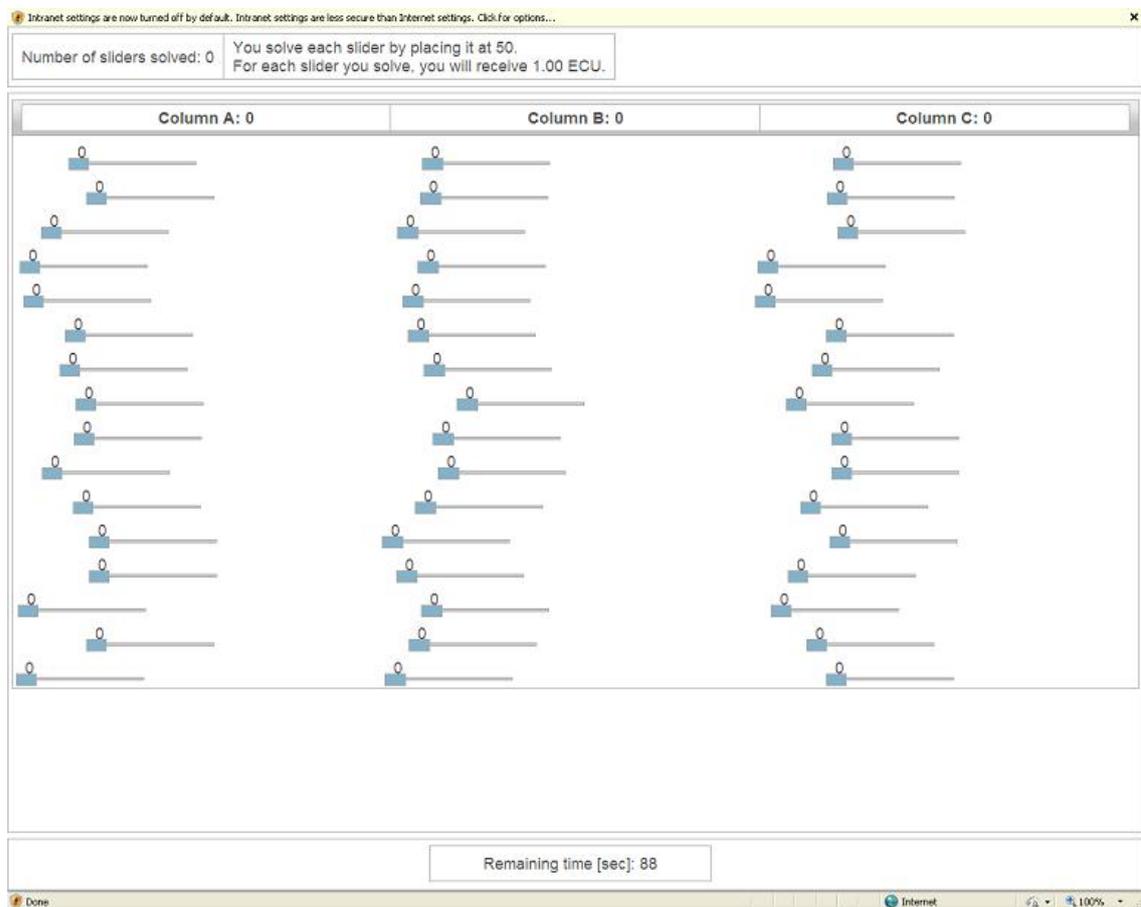
You will have two accounts, Account A and Account B. You must decide how much income to allocate to Account A and Account B. The two accounts differ in their returns.

For every 1 ECU you allocate to Account A:

- There is a 5 in 5 chance (or 100% probability) that you will lose 0.25 ECU.

For every 1 ECU you allocate to Account B:

- There is a 4 in 5 chance (or 80% probability) that you will lose 0 ECU.

- There is a 1 in 5 chance (or 20% probability) that you will lose 0.50 ECU.

You will need to fill a form, in which you state how much money you wish to allocate to Account A. The rest of your money will be allocated to Account B. Please click Next to see a picture of the form you must fill in.
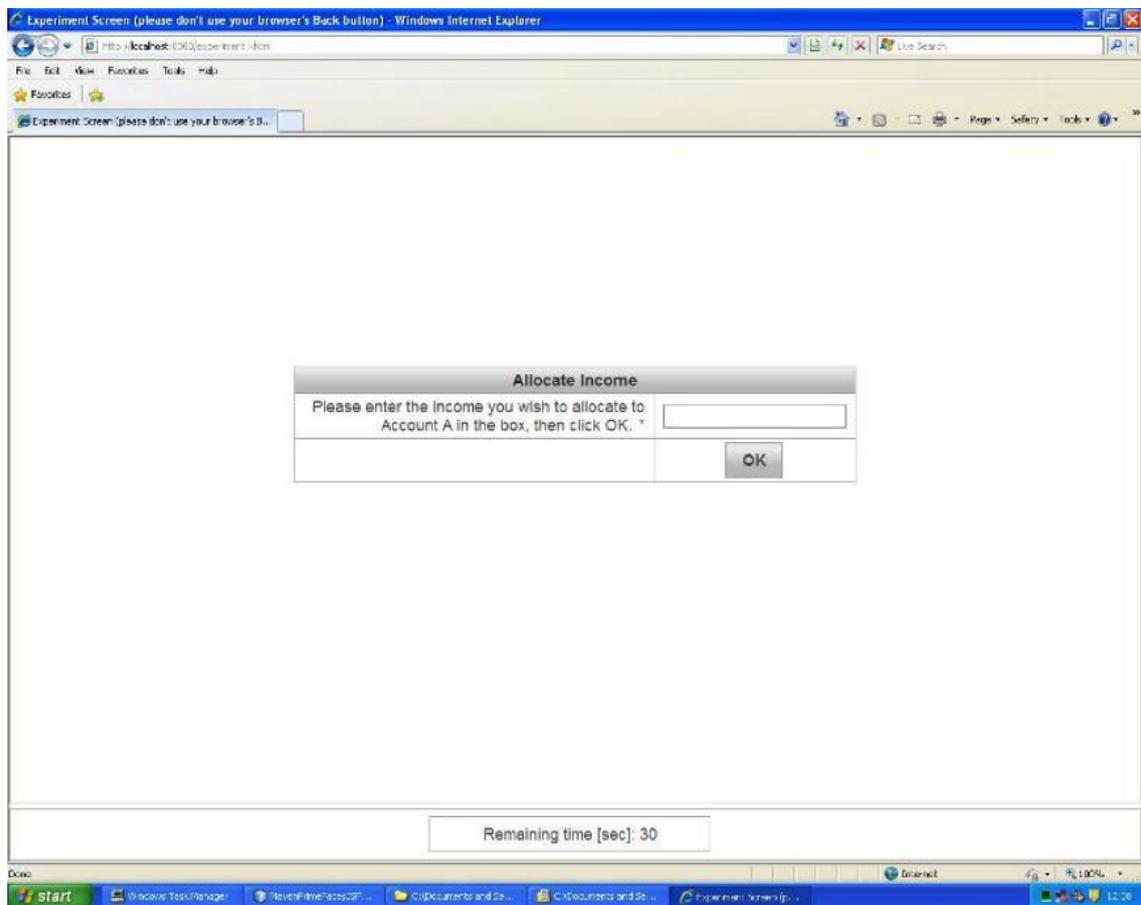
Now lets go through a few hypothetical examples.

Example 1:

Your Details	ECU
Your Income:	10.00 ECU
Income allocated to Account A:	10.00 ECU
(-) Your loss:	2.50 ECU
Payoff from Account A:	7.50 ECU
Income allocated to Account B	0 ECU
Loss incurred on Account B income?	Yes
(-) Loss (if any)	0 ECU
Payoff from Account B:	0 ECU
END OF ROUND INCOME	**7.50 ECU**

Example 2:

Your Details	ECU
Your Income:	20.00 ECU
Income allocated to Account A:	16.00 ECU
(-) Your loss:	4.00 ECU
Payoff from Account A:	12.00 ECU
Income allocated to Account B	4.00 ECU
Loss incurred on Account B income?	Yes
(-) Loss (if any)	2.00 ECU
Payoff from Account B:	2.00 ECU
END OF ROUND INCOME	**14.00 ECU**

Example 3:

Your Details	ECU
Your Income:	16.00 ECU
Income allocated to Account A:	10.00 ECU
(-) Your loss:	2.50 ECU
Payoff from Account A:	7.50 ECU
Income allocated to Account B	6.00 ECU
Loss incurred on Account B income?	No
(-) Loss (if any)	0.00 ECU
Payoff from Account B:	6.00 ECU
END OF ROUND INCOME	**13.50 ECU**

At the end of the round, you will be presented with a screen that summarizes the following:

- How much income you earned by solving sliders

- The amount of income you allocated to Account A and Account B

- Your final income 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 round.

Summary

In short, each round in this part of the experiment will consist of three stages:

STAGE 1: You earn income by solving sliders.

STAGE 2: You allocate your income to Account A and/or Account B.

STAGE 3: You get your final earnings, which depend on how you allocated your income.

This completes the description of Part A. The experiment will start automatically after 10 minutes, when the Remaining time [sec] hits zero.

A.6 Risk Elicitation Task Instructions

In this part of the experiment, we will ask you to make a number of decisions.

The decisions we will ask you to make will be choices between pairs of lotteries. The table below illustrates the task.

Each lottery can have two outcomes. The columns named Prob indicate the probability of each outcome of the draw, while the ECU columns indicate the value of the outcome.

For example, in Decision #1, picking Lottery A means you have a 10% chance of getting 40.00 ECU and a 90% chance of getting 32.00 ECU; picking Lottery B means you have a 10% chance of getting 77.00 ECU and a 90% chance of getting 2.00 ECU.

If you look carefully, as you move down the table the probability associated with the high payoff rises for both Lottery A and Lottery B.

In fact, once you get to Decision 10, you know that the high-value outcome for each lottery will happen with certainty. Your choice is therefore between 40 ECU for sure if you pick Lottery A or 77 ECU for sure if you pick Lottery B.

You must decide whether and when to switch from Lottery A to Lottery B.

If you wish to play Lottery A in all Decisions, type 11 in the box on the screen.

If you wish to play Lottery B in all Decisions, type 1 in the box.

Type a number between 2 and 10 in the box to switch to playing Lottery B starting with that Decision. For example, if you would like to play Decision A for Decisions 1 to 5 and B for Decisions 6 to 10, you should type 6 in the box.

After all participants in your group click OK, the computer will pick one of the ten Decisions at random and play out the lottery you chose for that Decision. The outcome of that lottery will determine your payoff for this part of the experiment.

B   Post-Experimental Survey Questionnaire

Q1. Where did you do the online experiment?

- Home

- Office

- Café or similar public space

Q2. Understanding the experiment. Please think about your understanding of the experiment when you were playing it. On a scale of 1 to 7, please tell us how much you agree or disagree with the following statements. Where 1 is 'strongly agree' and 7 is 'strongly disagree':

(a) I found it easy to remember my income when declaring how many sliders I'd completed.

(b) I understood that I could declare an income different to the amount actually earned.

(c) I understood that my declaration could be audited.

(d) I understood how likely I was to be audited.

(e) I understood that if I was audited and 'caught' under-declaring my income, I would be fined.

(f) I understood that I could potentially take more money home by under-declaring my income.

Q3. Motivation and behaviour during the experiment. Thinking about your general approach to the experiment, which of the following statements best describes how you behaved:

- I tried to always declare my income accurately.

- I occasionally declared less income than I had earned

- I mostly declared less income than I had earned

- I always declared less income than I had earned

Q4. Please explain in a few sentences why you took this approach to the experiment.

Q5. Please think about how you approached the experiment. On a scale of 1 to 7, please tell us how much you agree or disagree with the following statements. Where 1 is 'strongly agree' and 7 is 'strongly disagree':

(a) I declared all my income because it is the right thing to do.

(b) I declared all my income because evasion is unfair on others.

(c) I declared all my income because that was the rules.

(d) I didn't even think of not declaring all my income.

(e) I didn't put a lot of thought into the amount I declared, I just put what I'd earned.

(f) Not declaring all my income on the experiment would make me feel guilty.

(g) I was worried about being audited so I declared all my income.

(h) I was worried about being fined so I declared all my income.

(i) I treated the experiment like a on-line questionnaire and wanted to provide honest answers.

(j) I didn't see how I could cheat as the computer would know how much I earned from the task.

Q6. Please think about how you approached the experiment. On a scale of 1 to 7, please tell us how much you agree or disagree with the following statements. Where 1 is 'strongly agree' and 7 is 'strongly disagree':

(a) I took a calculated risk to not declare all my income.

(b) I wanted to earn as much from the experiment as possible.

(c) I had a target income to reach and under-declared to get to it.

(d) I under declared because I didn't think I would be audited.

(e) I under declared because I didn't think the fine was very high.

(f) I started to take more risks as the game wore on.

(g) I started to take less risks as the game wore on for fear of getting caught.

Q7. During the experiment I thought the likelihood of being audited was

- Lower than stated

- The same as stated

- Higher than stated

Q8. During the experiment I thought the likelihood of being audited was

- The same regardless of what I reported

- Higher if I did not report all my income

- Higher if I reported a low income

- Depended on when I was last audited

- Depended on the outcome of previous audits

- Other

- Don't Know