Deterrence Effects of Auditing Rules: An Experimental Study

Tan, Fangfang and Yim, Andrew

Tilburg University, CentER, Cass Business School, City University, London

17 September 2010

Deterrence Effects of Auditing Rules: An Experimental Study*

(Preliminary, please do not quote)

Fangfang Tan †
CentER,
Tilburg University

Andrew Yim ‡
Cass Business School,
City University London

September 17, 2010

Abstract

This paper examines the deterrence effect of two auditing rules via a laboratory experiment. A traditional rule which is usually assumed in the auditing literature, audits a taxpayer with a constant probability, which is independent of others' tax returns. A bounded rule recently proposed and analyzed in the literature chooses a sample from the population of reported low-income taxpayers to audit, taking into account the capacity of the auditor. We find that the deterrence effect of a bounded rule is as strong as that of a traditional rule, but is more cost-effective since fewer audits are conducted. The results lend further support to the bounded rule as a more cost-effective alternative to the traditional rule.

JEL Classification numbers: H26, M42, C9, C72

Keywords: Audit sampling plan, tax audit, tax compliance, tax evasion, experimental economics.

---

*We thank James Alm, Rudolf Avenhaus, Ralph-C Bayer, Jochen Bigus, Kim Bloomquist, Folkert Botma, Liesbeth Bruynseels, Patricio Dalton, Douglas DeJong, Marta Serra Garcia, Lata Gangadharan, Li Hao, Bastian Henze, Karla Johnstone, Steven Kachelmeier, William Kinney, Tobias Klein, Jan Potters, Ernesto Reuben, Jason Shachat, Joel Slemrod, Bin Srinidhi, Konrad Stahl, Jeroen Suijs, Eline van de Heiden, Arthur van Soest, Gijs van de Kuilen, Ann Vanstraelen, Eyal Winter, Shlomo Yitzhaki, Linke Zhu, and participants of the presentation at Tilburg University, 2010 Behavioral Decision Research in Management (BDRM) Conference held at Carnegie Mellon University and 66th Annual Congress of the International Institute of Public Finance for helpful comments and suggestions. We are particularly grateful for the valuable feedback provided by Luc Bissonnette, Eddy Cardinaels, Wieland Mueller, Charles Noussair, and Vilen Lipatov. We thank Chung Kim and Michael McKee for sharing their experimental instructions, and CentER lab for financial support. Siqi Pan and Huojun Sun provided excellent research assistance. All remaining errors are ours.

†Department of Economics, Tilburg University, Tilburg, the Netherlands. E-mail: F.tan@uvt.nl.

‡Cass Business School, City University London, 106 Bunhill Row, London, EC1Y 8TZ. E-mail: andrew.yim@aya.yale.edu.

1 Introduction

Auditing is universal in our society. For instance, external auditors examine company financial statement, corporate headquarters go through budget plans by business divisions, and regulators keep an eye on the operation of banks. To date, the existing auditing literature has focused on audit pricing, audit quality, and auditor independence (see, e.g. Nelson and Tan (2005)). However, the comparison of different auditing rules has arguably been under explored, given the potential for policy implication.

Tax auditing as an obvioius example.[1] The literature on tax compliance so far mostly assumes that audits are carried out in a simple randomized fashion, in which each taxpayer is independently selected for audit with a given probability (see, for example, Moser et al. (1995), Zimbelman and Waller (1999), Boylan and Sprinkle (2001), Kim et al. (2005), Kim and Waller (2005), and Alm et al. (2009)). We term this the *traditional* rule. One undesirable feature of the traditional rule is that audit resources are often used inefficiently. Imagine that a tax auditor is applying the traditional rule in examining a large number of tax return files. Since auditing decisions are carried on a random basis, the auditor must commit a budget which allows a full audit of all the files. Otherwise such strategy is not credible. However, these resources set aside for an audit unit's activities have an opportunity cost to the organization. For instance, in the fisical year 2005 and 2006, US IRS reserved $9,998 million and $10,461 million for tax administration and operations. The unused budget, however, were $80.6 million and $104.6 million respectively (US Department of the Treasury (2006)). If the IRS sets aside ample resources for audit purpose but could have provided the same deterrence with fewer resources, the resources could have been better used elsewhere. Consequently, an auditor might have greater latitude to formulate an audit strategy than simply conducting independent randomized audits. Such efficient auditing rules help to deter tax evaders in a cheaper way, and hence saves auditing resources.

One way to solve this problem is to formulate an auditing strategy on the basis of the entire taxpayer population. A recent model by Yim (2009) proposes and analyzes a new audit sampling rule named the *bounded rule*. Simply put, the bounded rule chooses an audit sample from the population of self-reported "low-income" taxpayers, given a pre-committed audit capacity. It audits

[1] Although we frame our entire analysis and experiment in the context of tax compliance, the application could extend to other related areas such as financial audits (i.e. with public accounting firms checking on company clients), internal audits (with corporate headquarters checking on business units), or other circumstances where regulators need to check on inspectees to enforce the compliance of regulations (e.g., Chen and Johnston (2008); Lennox and Pittman (2010)).

a random selected sample of self-reported "low-income" whenever the number of these reports exceeds audit capacity, or otherwise all of the reports. Yim (2009) shows that any compliance level induced by a traditional rule as a Nash equilibrium can also be induced by a bounded rule, but with substantially fewer reources committed.[2] Hence, research to understand further the properties of a bounded-rule has practical relevance to issues like audit staff planning and compliance enforcement efficiency.

As the first step to study the bounded rule, this paper aims at comparing its deterrence effect to the traditional rule in via a lab experiment. The laboratory allows us to directly test various auditing rules in a controlled environment. Such control enables us to isolate the many factors that confound behavior. In the meantime, it helps us to examine factors that are left out in the theory.

Our laboratory setting follows the basic features of the Yim (2009) model, which is an extension of the classic tax compliance game by Graetz et al. (1986). Every taxpayer has a certain probability of receiving high- or low-income. They have to decide simultaneously and independently whether to report their incomes truthfully to the auditor. The auditor deducts taxes according to players' reported incomes, and performs different auditing rules according to treatments.

In the original Yim (2009) paper, the auditor formulates an audit strategy based on his expectation of the taxpayers. In theory, we can construct the two auditing rules such that the induced compliance level is the same when auditors and taxpayers play Nash equilibrium in both games. Nevertheless, this requires a demanding understanding of the game and mutual belief towards each others' actions. Any off-equilibrium decisions by the auditors will lead to the behavior of the taxpayers uncomparable under the two rules. Hence, in this study, we control the auditor-taxpayer interactions by letting the auditor commits to an announced audit rule. In this way, we make sure that a bounded rule and a traditional rule induce the same deterrence power predicted by theory. Consequently, this paper should be not considered as a direct test of the Yim (2009) model. Instead, it focuses on examining whether and how human subjects react to the two rules.

We also examine the bounded rule under another parameter domain where the ex-ante probability of receiving high income increases. In reality, it resembles a rich neighborhood where every taxpayer is likely to earn a high-income. The game induced by the bounded rule has a payoff-

[2] In real life, the audit probability of the traditional rule, which is determined from the standard one-to-one analysis, has been interpreted as doing proportional sampling in settings with multiple inspectees. Simply put, it means randomly selecting a constant proportion of the suspicious inspectees (i.e., those filing "low-income" tax returns) for audit. Yim (2009) shows that the committed resource in the proportional sampling rule is still more than that of the bounded rule. In fact, the bounded rule is proved to be a cost minimizing rule to generate a given level of compliance.

dominant equilibrium in which all taxpayers underreport, and risk-dominant equilibrium in which all taxpayers report truthfully. We are interested in knowing whether the bounded rule still functions, which depends on the equilibrium subjects selected in the presence of multiple equilibria.

The main results are the following. The bounded rule induces the same compliance level as the traditional rule does. However, the bounded rule is more cost-effective for two reasons. First, it requires a lower amount of audits to sustain the same level of deterrence. Second, the committed resources are used more efficiently, i.e., the budget usage ratio is higher. In both treatments, the absolute level of compliance is higher than theory prediction. Such behavior can be best explained by a structural model incorporating loss aversion and stochastic measurement errors in forming utility. The bounded rule generates even higher deterrence in the presence of multiple equilibria, since taxpayers fail to coordinate on the zero-compliance outcome. All these results lend support to the bounded rule as a more cost-effective alternative to the traditional rule.

Our paper is related to two strands of literature. The first examines alternative auditing rules opposed to the simple random audit. For instance, Reinganum and Wilde (1985) analyze an "audit cuoff" policy in which an audit is triggered if the reported income is below a certain threshold, and otherwise no audit if the reported income is above the threshold. They show that there exists an equilibrium where the audit probability is decreasing in the level of reported income; and the all taxpayers under-report, although with an amount decreases in ture income. Bayer and Cowell (2009) build a model to examine the effect of a relative rule in a world where firms interact both on production and tax-compliance decisions. In their model, the tax authority first can commit a relative audit rule under which a higher audit probability is assigned to a lower reported-income. On the basis of this information, firms select their quantities and claim taxes. The result shows that the relative audit rule is more deterrent than a fixed, random audit rule when there is collusion among firms in either production or tax-declaration. The bounded rule in our study shares the key feature with the above rules that only low-income reports attract audit attention. Slemrod and Yitzhaki (2002) provides a detail discussion on these alternative auditing rules.

The second literature uses experiments as a tool to study auditing rules. The traditional rule and its variants are widely studied in this literature (see the literature review by Alm and McKee (1998)). A meta study by Blackwell (2007) based on twenty laboratory experimental studies finds that an increase in audit probability or fine rate leads to higher compliance, but the tax rate has no significant effect. Alm et al. (1993) examine a cut-off rule by combining a sure audit below a threshold on reported income and a small, random audit above the threshold. This cut-off rule, although is the most effective in increasing compliance, requires a large amount of audits. Alm and McKee (2004) consider another version of the cutoff rule with audit probability depending on the deviation of an individual's reported income from the average of the incomes reported by all other players. They find that it is difficult for players to coordinate to zero compliance without communication. With communication, players succeed in coordinating to extreme a low compliance level, but a cut-off rule combining some random audit solves this problem. Our setup differs from theirs in that the game induced by the bounded rule does not need to be a coordination game. Moreover, their focus is the cut-off rule itself and the situations under which it is more effective. Our study, on the other hand, aims at comparing the effects of two rules given a population with the same income distribution.

Our paper makes two contributions. First, it provides evidence for a new cost-saving way of doing audit selections. Although the literature only suggests that alternative audit rules contingent on strategic interactions among players might be more deterrent, nothing is unknown about the actual responses of taxpayers to these rules. We empirically show that the bounded rule is a more cost-effective than the traditional rule. Moreover, the bounded rule is robust in the presence of coordination, in that it is difficult for subjects to coordinate on the payoff-dominant outcome. Second, we perform structural estimation of non-expected utility for the first time using data from a tax compliance experiment. These behavioral models require less strict assumptions on cognitive reasoning or the ability to formulate correct beliefs on others, and hence provide a much more satisfactory account of behavior in our data. Moreover, the exercise of structural estimation allows the possibility of comparing alternative behavioral models.

For the rest of the paper, Section 2 describes the tax compliance model and auditing rules that are examined in the experiment. Section 3 presents the experimental design and procedures. Section 4 formulates the testing hypotheses. Section 5 analyses experimental data with both nonparametric and parameteric methods. Section 6 concludes and discusses directions for future research.

2 Model Description

The model used in this study follows the basic setup in Yim (2009), following the classic compliance game by Graetz et al. (1986). Consider a player population of size N. For simplicity, we assume there are only two income classes: high and low, denoted I_H and I_L, respectively, where $I_L < I_H$. Each player has a probability q of being a high-income taxpayer (H-type) and $1 - q$ of being a low-income taxpayer (L-type), where $0 < q < 1$. Players know the type distribution as well as their own types, but they do not know the exact types of the other players. Each player has to decide simultaneously and privately whether to report high-income (I_H) or low-income (I_L) to the tax authority. Let T_H and T_L be the tax payment by high- and low-income taxpayers respectively, where $T_H < I_H$, $T_L < I_L$, and $T_L < T_H$. If cheaters are audited, a fine F is imposed on top of the tax they should have paid. However, taxpayers who report truthfully are never fined and incur no cost if they are audited.

The *traditional rule* can be easily presented. Any taxpayer who has filed "low-income" reports will face a flat probability of a_{TR} being audited independently. Since reporting truthfully does not incur any cost by being audited, L-type players always state their income truthfully. If they report a high-income, they will be taxed T_H, which is strictly larger than the tax they need to pay if they honestly state income T_L. For H-type players, the honest-reporting payoff is $I_H - T_H$. If they underreport, the payoff is $I_H - T_L$ if they are not audited, and $I_H - T_H - F$ if they are audited. Therefore, they choose to underreport if the expected utility is strictly larger:

$$(1 - a_{TR})U(I_H - T_L) + a_{TR}U(I_H - T_H - F) > U(I_H - T_H)$$

where $U(\cdot)$ is the utility function. If the audit probability is less than the threshold a^*_{TR} defined by

$$a^*_{TR} = \frac{U(I_H - T_L) - U(I_H - T_H)}{U(I_H - T_L) - U(I_H - T_H - F)}$$

the H-type players are expect to underreport. Otherwise, if the audit probability is larger than a^*_{TR}, then they choose to report truthfully. In the case of risk-neutrality, this threshold becomes $\frac{T_H - T_L}{F + T_H - T_L}$.

Note, however, that the traditional rule does not model cost explicitly. In fact, it implicitly assumes that the tax agency has the budget to do a full audit of N files. Moreover, the expected number of audits is $a_{TR}N$. This means the larger number of "low-income" report turned in, the more files the auditor needs to check. In the following, we present an alternative auditing rule that takes into account the resources of the tax agency. It allows the tax agency to induce the same level of compliance with lower cost.

The *bounded rule* requires the auditor to first set a committed budget characterized by the maximum number of K audits allowed. It then constructs an audit sample size contingent on the number of "low-income" reports L. If L is smaller or equal to the audit capacity K, the auditor will audit all L reports. However, if L is strictly larger than K, then the auditor will randomly audit K reports. To put it more formally, the bounded rule selects an audit sample size $s(L) = \min\{K, L\}$. Every "low-income" taxpayer faces a probability $a_L = \min\{K/L, 1\}$ of being audited (for $L = 0, 1, ...N$).

To illustrate how the bounded rule works, we focus on the analysis of H-type players, as L-type players again have a dominant strategy of reporting truthfully.[3] Similar to the traditional rule, the H-type players face the tax evasion gamble of choosing a sure payoff of $I_H - T_H$, or a high payoff of $I_H - T_L$ if they are not audited but a low payoff $I_H - T_H - F$ otherwise. Unlike the traditional rule, however, the audit probability, denoted by a_{BD}, is no longer an exogenously given. Instead, it depends on player i's subjective belief on the likelihood of the proportion of "low-income" reports turned in by other players, denoted by B_i.

A "low-income" report could come from two sources. The first source is from a truth-telling L-type player with probability $1 - q$. Alternatively, it could come from H-type players lying as L-type. If player i thinks that the under-reporting probability of H-type players is b_i, this scenario will occur with probability qb_i. Hence the overall probability of observing a "low-income" report B_i should be the sum of the probabilities in these two situations: $B_i = 1 - q + qb_i$. The overall audit probability perceived for lying H-type taxpayers is now the sum of probabilities given that none, one, two, ... or all $N - 1$ other players submit "low-income" reports, or more formally,

$$a_{BD} = \sum_{n=0}^{N-1} a_L \mathbf{Bin}(n, N-1; B_i)$$

where $\mathbf{Bin}(n, N-1, 1-q+qb_i)$ is the probability that exactly n out of $N - 1$ players submit "low-income" reports. **Bin** represents the binominal distribution.

We assume that players are homogeneous, individual profit maximizers, and their beliefs are symmetric in the equilibrium. Consequently, for a given set of parameters N, K, q, the game among taxpayers induced by the bounded rule always exists an equilibrium.

[3] The actual percentage of honest reports among L-type taxpayers are 99.68% and 99.28% across treatments, suggesting that they do play the dominant strategy.

3 Experimental Design and Procedure

We design our experiments to compare the deterrence effects of two auditing rules, and to study players behavior under these mechanisms. The idea is to choose parameters which induces the same level of compliance in both treatments. Based on the capacity constraint in the lab, we fix the size of the taxpayer population $N = 8$, and the audit capacity constraint $K = 2$ for the bounded rule..

The tax compliance game in both treatments has three stages: income reporting/tax deduction, audit/fine deduction, and feedback. Subjects are endowed with either a high income I_H of €25 or a low income I_L of €10 with probability $q = 0.5$. Subjects are informed about the group size and the income distribution. During the income-reporting stage, they have to decide simultaneously and privately the type of income to report to an auditor, which is simulated by a computer. The computer automatically deducts taxes according to the reported incomes. The tax for subjects reporting "high income" (T_H) is €12.5, whereas the tax for subjects reporting "low income" (T_L) is €2.5.[4] Subjects are told that taxes are deducted based on their reported incomes instead of true incomes. For instance, high-income taxpayers will receive €22.5, instead of €12.5, if they submit "low-income" reports. Similarly, low-income taxpayers will receive -€2.5, instead of €7.5, if they submit "high-income" reports.[5] In the audit stage, subjects reported €25 are never audited. Subjects reported €10 potentially are subject to an audit depending on the treatments.

Traditional: We use a traditional rule as our baseline treatment. Subjects filing "low-income" reports face an independent audit probability of 0.4. This audit probability induces comparable compliance rate to the bounded rule. If they indeed report honestly, nothing will happen with respect to their final payoffs. However, if cheaters are caught by the auditor, they need to pay back the €10 of tax evaded plus a fine F of €10.[6]

Bounded: This is our target treatment. The audit probability depends on the total number of "low-income" reports received. The maximum number of audits to be conducted is $K = 2$. This means if the number of low-income reports does not exceed two, all of them will be audited with probability 100%. Otherwise, the audit probability is monotonically decreasing on the number of

[4] We attempt to choose the experimental parameters concerning taxation in line with the reality. For instance, the real-world tax rates for high-income and low-income taxpayers are usually dependent on the levels of their incomes. In particular, many European countries such as Britain, Germany, Italy and the Netherlands use a progressive tax system instead of a proportional one. Since we conducted this experiment in Europe, we decided to adopt a simplified version of a progressive tax system for the sake of facilitating subjects' understanding.

[5] Even when a subject with low income makes a loss by submitting "high income" reports and later this decision was selected for payment, the potential loss will be still be covered by the show-up fee. During the experiment sessions, this situation never actually happened.

[6] We implicitly assume that the auditor has sufficient resources in place to implement the traditional rule, which potentially requires auditing all eight subjects in the taxpayer population of our experiment.

low-income reports L. In particular, the probability is 0.67 for $L = 3$, 0.5 for $L = 4$, 0.4 for $L = 5$, 0.33 for $L = 6$, 0.29 for $L = 7$, and 0.25 for $L = 8$. This K parameter guarantees a unique Nash equilibrium based on non-cooperative game theory. Fine for cheaters is exactly the same as the *Traditional* treatment.

The experiment was conducted at the CentER Lab in Tilburg University from October to December 2009. Tilburg University students, mostly major in economics or business, participated as subjects in the experiment. Each treatment consisted of four sessions of 16 subjects each. The duration of a session was about 1 hour (including the initial instruction and final payment to subjects). The average earnings are €16.23. We used Z-Tree software to program and conduct the experiment (Fischbacher (2007)).

The instructions used in our tax compliance game were modified from instructions in prior studies of the literature, namely Alm et al. (2009), Kim et al. (2005), and Kim and Waller (2005) (see Appendix B.2). At the beginning of each session, subjects were randomly assigned to the computer terminals. After finishing reading instruction, we had some quiz questions to test the subjects' understanding of the game. Only after all subjects had answered all the questions correctly did we start the experiment.

The game had 30 periods of play. At the beginning of each period, we randomly allocated 16 subjects into two groups of eight each to guarantee that a subject will not play with the same group of participants again. This random re-matching design simulates a one-shot scenario but allows the subjects to be familiar with the game environment. Since the group composition kept changing from period to period, collusion among the subjects was extremely difficult. At the end of each period, a summary screen was shown to subjects with feedback information. The feedback information included: (i) the subject's true income and the income reported and (ii) the final payoff for the period, including a fine, if any, for tax evasion. Subjects were not informed of others' payoffs.[7]

After completing the tax compliance experiment, subjects were asked to finish a risk elicitation task similar to the one used by Holt and Laury (2002).[8] The risk elicitation task required subjects to make selections of a set of 21 lottery pairs. The purpose of this task is to obtain individual risk aversion level and later use it to explain behavior in the tax experiment.

At the end of the experiment, the subjects were asked to fill out two questionnaires. The first

[7] We do not inform the subjects whether they are audited. The purpose of not explicitly present such feedback is to discourage players experiencing the audit probability when they are L-type. As for the H-type players, it is very easy to know whether their file have been audited or not from the final payoffs.

[8] We handed out the instructions for the risk elicitation task only after the tax compliance game, so that the subjects were not aware of its existence beforehand.

one concerns social background information such as gender, nationality, and experience of learning economics. Moreover, we also elicited subjects' Machiavelli scores by means of the Machiavellian scale personality test (see Christie and Geis (1970)).[9] The second one contains questions on their perceptions of the treatment designs in order to assess the effectiveness of treatment manipulations.

Upon finishing the questionnaires, we randomly selected one period of the tax game and the realization of one lottery to determine the final payment of a subject. This random payment scheme mitigates the potential income effect that could confound experimental results.[10]A show-up fee of €3 was added on top of the total earnings so that no subjects left the lab with loss.

4 Testing Hypotheses

We present our hypotheses regarding the deterrence effects of both rules. The deterrence effect is indicated by the underreporting rate in the population, namely, the proportion of high-income taxpayers filing "low-income" reports in a certain period. As discussed in Section 2, our analysis will focus on the H-type taxpayers, as the L-type players have a dominant strategy of reporting honestly regardless of the auditing rules.

In the following analysis, we let h be the honestly reporting strategy for H-type players, and u to be the underreporting strategy. As the audit probability a_{TR} is set to be 0.4 for the traditional rule, an under-reporting decision is equivalent to selecting a lottery of €22.5 with probability 0.6 and €2.5 with probability 0.4. The expected payoff is therefore: $E(\pi_u) = €22.5 \times 0.6 + €2.5 \times 0.4 = €14.5$. As it is strictly larger than the sure payoff €12.5 of reporting honestly, H-type taxpayers are expected to underreport.

Under the bounded rule, the audit probability is not exogenously given, but depends on perception of the players. Recall that the probability of n "low-income" reports submitted by the remaining $N-1$ taxpayers follows the binomial distribution $\mathbf{Bin}\ (n, N-1, B_i)$. If a H-type player decides to underreport, the sample size of "low-income" report will increase by 1 and every reported "low-income" player faces an audit sample size $s(1+n;K)$, with correspondence audit probability of $a_L = \min\{K/(1+n), 1\}$. In summary, the expected payoff of lying H-type taxpayers is now the sum of the payoffs given that none, one, two..., or all $N-1$ of other players handing in "low-income" reports:

[9] This test measures a person's preposition to be opportunistic and manipulative; with higher scores meaning that these properties are more pronounced.

[10] The Starmer and Sugden (1991) study shows that such a random lottery incentive system does not distort a subject's true preference.

$$E(\pi_u; N, q, K) = \sum_{n=0}^{N-1} \mathbf{Bin}(n; N-1, B_i)\left[(1-a_L)\times(I_H - T_L) + a_L \times (I_H - T_H - F)\right]$$

The Nash equilibrium of this game can be reached by elimination of dominated strategies. Reporting high-income is a dominated strategy for L-type players, since they will be taxed accordingly and incur a strictly lower payoff than otherwise. If the H-type players believe L-type obey dominance, then the strategy of reporting truthfully (h) is dominated. Assuming symmetry and unbiased belief among players, we can derive the equilibrium under-reporting decisions.

Proposition 1 *The game introduced by the bounded rule is dominance solvable. In this equilibrium, both the L-type and H-type players report low-income.*

Proof: See appendix A. ∎

Note that the above hypothesis holds for strategic, self-interest profit maximizers. Now suppose that some players are intrinsically honest, namely, they report their income truthfully regarless of their type. This assumption does not change the direction in terms of treatment differences. Recall that in the *Bounded* treatment, the optimal strategy of the H-type players does not depend on their beliefs towards other H-type players. Why? As long as they believe that L-type will not play dominated strategy (i.e. reporting high-income), they can form expectation on the proportion of "low-income" reports filed in each realized income distributions. Since the ex-ante probability of being a L-type player is sufficiently high ($q = 0.5$), a H-type player finds that the sure payoff of reporting honestly to be lower than then expected payoff from underreporting, even when s/she does not expect any other H-type to underreport. The analysis in the *Traditional* treatment is simpler. As player decisions are independent, the presence of honest players will not affect decisions of the self-interest profit maximizers. If the percentage of intrincially honest players is the same in both treatments, the compliance rate should be the same. For the mathematical formulation see Appendix A.

We build our hypothesis based on the above proposition:

Hypothesis 1 *The under-reporting rate is the same under both rules:* $b^{TR} = b^{BD}$.

The expected number of audits under the *Tradtional* treatment, L^{TR}, depends on the number of "low-income" reports. Let p^{TR} denote the percentage of players submitted "low-income". If

$p^{TR} > \frac{5}{8}$, L^{TR} will be larger than that 2 audits committed in the *Bounded* treatment. If the cost of an audit is the same in both treatments, fewer audits in the *Bounded* treatment means lower implementation cost. Hence, we advance the following hypothesis:

Hypothesis 2 *If the percentage of "low-income" report submitted is larger than 62.5% , the number of audits is smaller in the Bounded treatment than in the Traditional treatment. That means, the bounded rule is cheaper to implement to achieve the same level of compliance.*

5 Results

5.1 Average Treatment Effect

Table 1 summarizes descriptive results of non-compliance behavior and profit across experimental treatments. All statistics reported in this table are on session level. The left panel contains averages over all 30 periods of play, and the right panel is the results for the last 10 periods, where behavioral pattern is more stable.

Table 1: Summary statistics accross treatments(standard errors in parentheses)

	All 30 Periods		Last 10 Periods	
Overall Characterization	*Traditional*	*Bounded*	*Traditional*	*Bounded*
All subjects				
High-income probability	0.514	0.491	0.527	0.519
	(0.007)	(0.039)	(0.042)	(0.038)
Percentage of “low-income” reports	79.741%	78.853%	77.969%	75.935%
	(0.074)	(0.015)	(0.066)	(0.018)
H-type subjects				
Under-report frequency	60.829%	57.114%	58.163%	53.321%
	(0.144)	(0.049)	(0.143)	(0.052)
Average under-report profit	14.513	16.446	13.272	15.781
	(0.650)	(0.285)	(0.967)	(1.979)
Auditing statstics				
Total audit number	153.751	120	53.75	40
	(18.140)	(0.000)	(8.098)	(0.000)
Average audit number	2.563	2	2.692	2
(per group per period)	(0.300)	(0.000)	(0.414)	(0.000)
Audit frequency	40.161%	31.712%	42.958%	32.943%
	(0.030)	(0.006)	(0.038)	(0.007)
Budget usage ratio	32.033%	100%	32.091%	100%
	(0.181)	(0.000)	(0.181)	(0.000)
Cheater detection rate	38.762%	33.134%	42.082%	31.876%
	(0.065)	(0.043)	(0.107)	(0.125)

The two rows on top of the table report statistics concerning all subjects. The first row indicates that the actual probability of being a H-type in both treatments is very close to their pre-specified levels with repeated drawing. The second row displays the percentage of low-income reports among all reports, including reports from L-type taxpayers and the fake ones by H-type players. This number is around 80% in both treatments, which satisfies the condition in Hypothesis 2 that allowing us to compare the implementation costs of two rules.

The next two rows focus on H-type taxpayers, whom are the main interest of our study. The third row reports the overall under-report frequency, which is 60.83% in the *Traditional* treatment and 57.11% in the *Bounded* treatment. A two-sided Wilcoxon rank-sum test cannot reject the null hypothesis that the under-report frequency of the two treatments are the same ($p = 0.386$), although in the last 10 periods, the difference becomes a bit larger ($p = 0.193$). The profit for the cheaters in the *Bounded* treatment is €2 higher ($p < 0.05$), maybe due to the fact that the detection rate in this treatment is lower than that in the traditional rule (see the last row).

The final four rows concern audit statistics. We find two pieces of evidence supporting the bounded rule to be more cost-effective. To begin with, it can sustain the same level of compliance with a lower cost. Due to the fact that the audtior effectively commits to fewer audit resources under the bounded rule, both the total audit number and the audit frequency are significantly lower ($p < 0.05$). This result is pretty robust even when we compare this average number of audits per group per period ($p < 0.05$).

Apart from a lower implemenation cost, the bounded rule has a higher budget usage ratio. We define budget usage ratio to be the percentage of resources actually used for a given committed budget. This figure is 100% in the *Bounded* treatment, which means that all resources committed are used at their full capacity in each period (i.e. two audits). Under the traditional rule, the budget usage ratio is only 32.03%. The inefficiency comes from the fact that while auditor has to prepare resources to do all eight audits in each period, only a small fraction of audits are actually carried out.

The last column shows the effectiveness in cheater detection. The success rate is higher in the *Traditional* than the *Bounded* treatment, though not statistically significant ($p = 0.103$).

By comparing behavior over 30 periods (left panel) to behavior in the last 10 periods (right panel), we can see that in both treatments, the under-report rates decrease over time. Due to fewer "low-income" reports, the relative audit frequencies increase, and cheaters earn less. Nevertheless, the results of cross treatment comparisons remain the same. We summarize two results in this section:

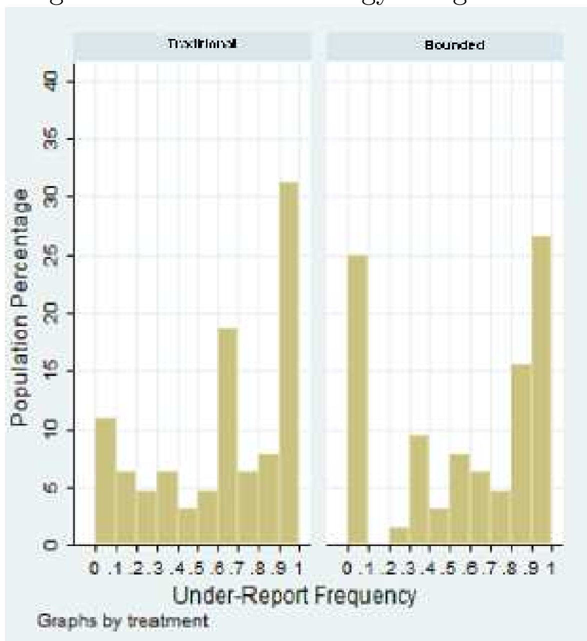
Result 1 *Hypothesis 1 is supported. The observed under-reporting rates are statistically the same in both treatments, although the absolute levels are significantly lower.*

Result 2 *Hypothesis 2 is supported. The bounded rule uses resources more efficiently in that 1) The average number of audits in the Bounded treatment is significantly smaller than that in the Traditional treatment, and 2) The budget usage ratio is higher.*

5.2 Individual Behavior in the Game

Figure 1 displays the distribution of individual strategies for H-type players across treatments. The horizontal axis represents subjects' under-reporting frequency throughout the game, i.e. the percentage of times when they receive high-income and decide to under-report. The vertical axis represents the percentages of players having the same strategy in each treatment.

Figure 1: Individual strategy categorization



A Mann-Whitney ranksum test cannot reject the null hypothesis that the strategy distributions in the *Traditional* and *Bounded* treatments are the same ($p = 0.3224$). According to this figure, only 28.13% of the subjects in the *Traditional* treatment and 20.31% of subjects in the *Bounded* treatment behave exactly in accordance with theory, namely, underreport when they are H-type throughout the experiment. On the other hand, the percentage of intrinsically honest subjects is 12.5% and 15.63% respectively. In sum, we can only explain about 40% of the data in the *Traditional* treatment and 35% in the *Bounded* treatment. The remaining subjects report their income truthfully even though evasion is attractive, at least to expected payoff maximizers. However, their behavior are quite stochastic, in that they switch between the two options with various frequency.

How could we bridge the gap between theory and our data in the experiment? To begin with, we argue that theory based on individual profit maximization makes two unrealistic assumptions on behavior. The first one is the assumption of perfect rationality. In reality, however, people are usually bounded by the cognitive limitation of their minds given the amount of time they have to make decisions. The second one is risk neutrality. Experimental literature documents mounting evidence that subjects are not risk neutral profit maximizers, but risk averse utility maximizers.

To relax the assumption of perfect rationality, we propose discrete choice model as a framework to accommodates boundedly rational behavior (McFadden (2001)). Models under such framework are motivated by empirical studies in which observed decisions exhibit some noises (see, e.g. Fischbacher and Stefani (2007), Loomes (2005), Rieskamp (2008) and Wilcox (2009)). These noises could come from observed sources like decisions errors, but also come from other unobserved or modeled channels such as individual perception of the game, or the sensitivity to payoff changes. Due to the presence of these noises, people make decision errors and hence are not consistent with their choices. Our *Baseline* treatment is essentially a non-strategic choice-under-uncertainty problem to H-type taxpayers. Therefore, the classic individual discrete choice model is a natural setting to explore behavioral anomalies. The bounded treatment introduces interactions of players. A general way to incorporate decision error is quantal response equilibrium first proposed by McKelvey and Palfrey (1995), which is based on the random utility maximization model of McFadden (1973).

According to the discrete choice framework, H-type taxpayers will choose to under-report if and only if the difference in the expected utilities is sufficiently large to exceed a stochastic error denoted by ε, i.e.,

$$EU(\pi_l) - \pi_h > \varepsilon$$

where ε is commonly assumed to be independently and identically distributed across players and actions with a Type 1 extreme value ("logit") distribution. The error can come from many sources such as the inability to calculate the expected payoff or trembling hands during decision making. A standard result of the discrete-choice model framework is that under the above error distributional assumptions, the choice probability b for lying, i.e., the underreporting probability, is given by the relation below:

$$\begin{aligned} b &= \Pr\{P(l) = 1\} \\ &= \Pr\{EU(\pi_l) - \pi_h > \mu\varepsilon\} \\ &= \frac{1}{1 + \exp\left[-\frac{EU(\pi_l)-\pi_h}{\mu}\right]} \end{aligned} \tag{1}$$

The parameter $\mu > 0$ captures the sensitivity of subjects' choices to the relative payoffs of the two choices. When μ approaches infinity, players choose under-reporting and honest-reporting with equal probability, independent of the relative expected payoffs. When μ decreases, on the other hand, players put less probability weight on choices that yield suboptimal payoffs, and the probability that they make the optimal choice converges to 1 when μ approaches to 0. Put it differently, μ is an index of the measurement error when subjects form expected utility from under-reporting. This is particularly true for subjects in the *Bounded* treatment, since the expected payoff from under-reporting is computationally demanding. Even for subjects in the *Traditional* treatment, more than half of the subjects switch between options due to unobserved reasons.

Within this framework, we can further relax the assumption of risk neutrality. In particular, we estimate and compare three behavioral models: risk-aversion, loss aversion with and without combining probability weighting. In the risk-aversion model, we assume that subjects have a CRRA-form utility function $u(\pi) = \frac{\pi^{1-r}}{1-r}$.[11],[12]. This model also allows us to explicitly test the assumption of risk neutrality. If the estimated r is significantly different from zero, we can reject the null hypothesis that subjects are risk neutral.

While the observed compliance behavior can be explained by risk attitude, it is also consistent with the notion of loss aversion. Recent research has shown that that loss aversion provides a much better account of tax evasion both in the lab and the field (see, e.g. , Elffers and Hessing (1997); Yaniv (1999); King and Sheffrin (2002); Dhami and al Nowaihi (2007); Dhami and Al-Nowaihi (2010)). The loss aversion model characterizes individuals as loss averse in terms of reference income, denoted by R. For a given amount of money, $x > 0$, and the value function $v(x)$ (specified below), losses are weighted more than gains ($|-v(-x)| > v(x)$). We follow Dhami and al-Nowaihi (2007,2010) by taking the honest ex-post tax income as the reference point: $R = I_H - T_H$. The rationale of this reference point is as follows. If the reference point is selected differently, say the initial income or the income after cheating detection, then taxpayers are always in the domain of losses or gains. In those cases, the asymmetry of gains and losses disappear, and we completely fall

[11] Alternative utility forms such as CARA and power-expo utility do not change the fit of data.

[12] Data from tax compliance game alone does not have any identification power to jointly three parameters, since it only contains two moments (i.e. the fraction subjects selecting the "risky" lottery in the traditional rule and that in the bounded rule) given a fixed payoff structure. To gain enough identification power, we pool data from both risk elicitation task and tax compliance game.

back to the framework of expected utility [13]. The income relative to the reference point is:

$$\pi_i = \begin{cases} I_H - T_H - F - R & \text{for} \quad i \text{ is caught} \\ I_H - T_L - R & \text{for} \quad i \text{ is not caught} \end{cases}$$

The form of the utility function follows Tversky and Kahneman (1992) defined separately over gains and losses: $U(\pi_i) = \pi_i^{\alpha}$ if $\pi_i \geq 0$, and $U(\pi_i) = -\lambda(-\pi_i)^{\beta}$ if $\pi_i < 0$. The α and β are the parameters controlling for the curvature of the utility functions, and λ is the coefficient of loss aversion. Subjects are considered loss-averse if $\lambda > 1$.

Besides value functions, subjects could also have a nonlinear transformation of the probability scale, i.e. they overweight low probabilities and underweight high probabilities (see, e.g. Kahneman and Tversky (1979)). In order to examine the effect of subjective probability weight, we estimate a third model combining loss-averse utility form with probability weighting function. In particular, we adopt a popular form of one-parameter probability weighting function: $w(p) = \frac{p^{\delta}}{(p^{\delta}+(1-p)^{\delta})}$, where $\delta \geq 0$. Note that if $\delta < 1$, the weighting function has an inverted "S" shape, which is concave for low probabilities and convex for high probabilities, and crosses the diagonal at the probability about 1/3.

Recall that H-type players are choosing between a safe lottery and a risky one with fixed probabilities in the traditional rule, but endogenous probabilities under the bounded rule. In the following, we let parameter "a" representing the perceived audit probability in the *Bounded* treatment. We are interested in the following question: If we transform a bounded rule into the context of a traditional rule, which exogenous audit probability "a" best justifies behavior? Moreover, how do risk attitude, probability weight, or loss aversion influence subjects' perception of the audit probability? The likelihood of under-reporting responses, conditional on the analytical method being true, depends on observed choices in different treatments. The conditional log-likelihood is the following:

$$\begin{aligned} \ln L(a_{it}|y_{it}) &= \sum_{i,t} \left\{ y_{it} \cdot \ln \left(\frac{1}{1+\exp[\frac{\pi_h - E(\pi_{lie})}{\mu}]} \right) + (1-y_{it}) \ln \left(1 - \frac{1}{1+\exp[\frac{\pi_h - E(\pi_{lie})}{\mu}]} \right) \right\} \quad (2) \\ E(\pi_{lie}) &= \begin{cases} 0.6 \times 22.5 + 0.4 \times 2.5 & \text{for} \quad i \in \textit{Traditional} \\ (1-a) \times 22.5 + a \times 2.5 & \text{for} \quad i \in \textit{Bounded} \end{cases} \end{aligned}$$

[13] To be more specific, such framework is called Rank dependent expected utility theory (RDEU), which can be considered as expected theory applied with a transformed cumulative probability distribution. See Dhami and al-Nowaihi (2007) for more detail.

where $y_{i,t} = 1(0)$ denotes that the subject i under-report (honest-report) in the tax compliance game in period t. The results are reported in Table 2.

Table 2: Comparison of behavioral models

	Risk aversion		**Loss aversion**		**Loss aversion & Probability Weighting**	
	Traditional	*Bounded*	*Traditional*	*Bounded*	*Traditional*	*Bounded*
Risk magnitude r	0.366***	0.594***				
	(0.350)	(0.055)				
Gain domain curvature α			0.445***	0.428***	0.640***	0.533***
			(0.034)	(0.038)	(0.459)	(0.075)
Loss domain curvature β			0.548***	0.708***	0.586***	0.858***
			(0.052)	(0.030)	(0.068)	(0.073)
Loss aversion coefficient λ			1.100***	1.148***	1.674***	1.283***
			(0.802)	(0.030)	(0.123)	(0.171)
Weighting parameter δ					1.150***	0.899***
					(0.193)	(0.120)
Perceived audit prob. a		0.336***		0.305***		0.240***
		(0.017)		(0.007)		(0.023)
Noise μ	0.667***	0.618***	0.266***	0.256***	0.430***	0.289***
	(0.067)	(0.098)	(0.019)	(0.424)	(0.042)	(0.040)
Log-likelihood	-1163.773	-1087.292	-1141.710	-1082.473	-1141.353	-1082.111
Observations	2331	2287	2331	2287	2331	2287

Notes: *10% significance; **5% significance, ***1% significance. The standard errors are clustered on subjects.

At the first glance, all parameters in these models are significant, suggesting that the alternative behavioral models help explaining the compliance behavior in our study. For instance, the risk aversion specification tells us that subjects are risk averse in both treatments, as the CRRA coefficient r is significantly larger than zero. The indicates risk aversion helps to explain our data. The perceived audit probability for a risk averse subject in the *Bounded* treatment is about 0.34. The explanation is straight-forward: To induce a similar compliance pattern among subjects who are risk-averse, we only need to "tune" the audit capacity of the bounded rule such that it is equivalent to a traditional rule with audit probability $a = 0.336$. In other words, we need fewer resources to acheive the same level of compliance for risk-averse subjects than risk-neutral ones.

In the loss aversion specification, subjects in both treatments exhibit loss aversion: The coefficients of the loss aversion parameter λ are larger than 1 in both treatments, which means subjects are more sensitive to loss than comparable magnitude of gain. The slopes of the value function indicate concavity in the gain domain (α) and convexity in the loss domain (β). Moreoever, A Vuong test on non-nested models favors the loss aversion model over risk aversion model ($z = -3.690$). If subjects are loss-averse, the bounded rule is even cheaper to implement, as we only need to induce the compliance rate similar to a traditional rule with audit probability $a = 0.306$.

We also run a third model combining loss aversion utility and probability weighting together. However, we do not find significant improvement in the likelihood. Moreover, the probability weighting parameter δ is not significantly different from 1 for both treatments ($p = 0.438$ and 0.397 respectively). It means that the average subjective probability of the subjects is pretty much in line with the objective audit probability. Overall, the results seem to indicate that the driving force for behavior is more likely in the way they view losses and gains, rather than the way they assess probabilities.

Figure 2: Observed and predicted non-compliance rates

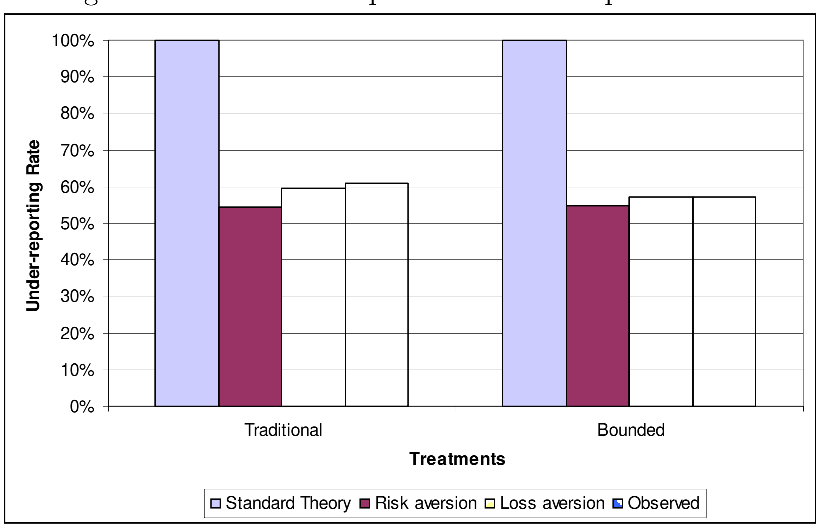


Figure 2 displays the observed and predicted under-reporting rates based on risk and loss aversion models. Since estimation results suggest that probability weighting does not explain the data well, we take parameter estimates from the second specification: loss aversion without probability weighting. Among the three models, the one by loss aversion fits our data the best. Result 3 summarizes the section.

Result 3 *The proportion of compliance behavior in both treatments is consistent with the presence* *of loss aversion together with some stochastic decision errors, although not in probability weighting.*

5.3 Coordination under the Bounded Rule

So far, the game introduced by the bounded rule is dominance solvable. In fact, it is not difficult to show that as long as the ex-ante probability of receiving high income is lower than 0.5, the H-type players always under-report given the dominant strategy of L-type. Essentially, the more L-type players who for sure honestly state their type in the population, it easier for the H-type players to pretend to be L-type.

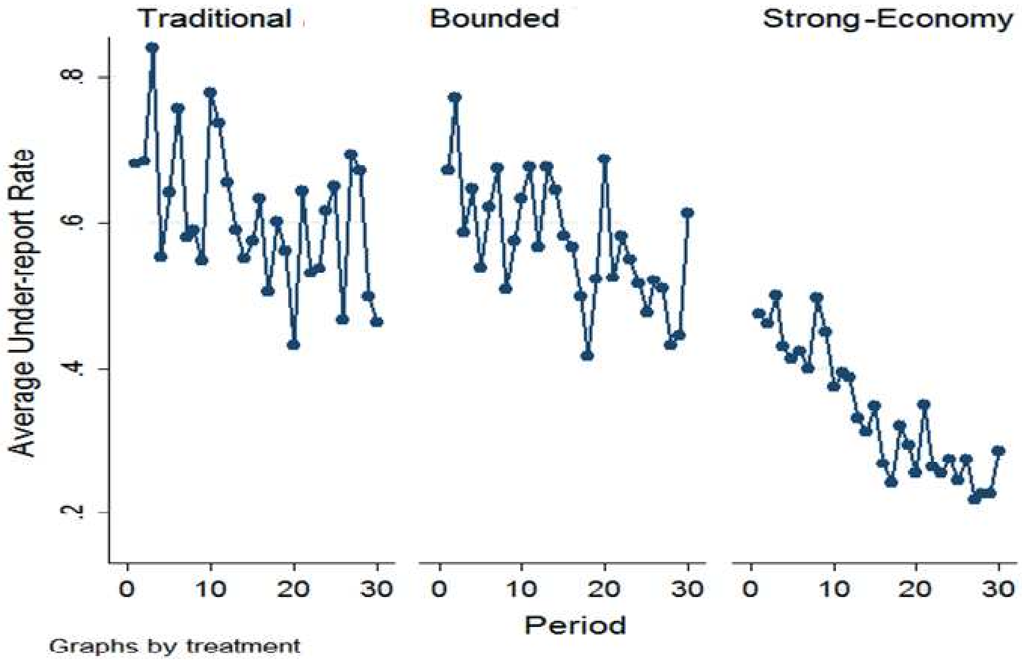
In this subsection, we examine the bounded rule in another paramter domain where the game has multiple equilibria. In this new bounded-rule treatment called *Strong-Economy*, everything remains the same as *Bounded* treatment except that the ex-ante probability of receiving high-income q becomes 0.9 instead of 0.5. According to standard game theory, the introduction of the bounded rule with the same audit capacity changes the interaction of players into a coordination game with incomplete information. There are two pure strategy NEs and one mixed strategy NE in the game. In the pure strategy equilibria, L-type players play their dominant strategy of reporting truthfully. All H-type players play under-reporting (truth-reporting) if they believe other H-type players are going to cheat with probability higher(lower) than 0.432. There is also a symmetric mixed strategy NE $b^{SE} = 0.432$.[14]

Studying behavior under a different q parameter has practical purpose as well. Remember the discussion in section 2 that the parameter q determines an important property of the bounded rule: A high q resembles a rich neighborhood where each inhabitant is very likely to be wealthy. Hence, this new treatment helps us explore the performance of the bounded rule at the high end of the income distribution. In particular, we are interested in knowing whether the bounded rule loses the deterrence effect in the presence of coordination.

The overall under-report rate in the *Strong-Economy* treatment is 33.95% over all 30 periods, and 26.16% in the last 10 periods. According to Figure 3, it is clear that the deterrence effect is the strongest, as the under-report frequency is significantly lower compared to the other two treatments (a two-sided Mann-Whitney ranksum test with $p < 0.05$). This difference is already salient in the first period, and remains highly significant throughout the game. Interestingly, although the total number of audits is smaller in this treatment even compared to the *Bounded* treatment ($p < 0.05$),

[14] For the proof please refer to appendix A. Note that there are other asymmetric equilibria in the game. However, we ignore them in a symmetric setting, since these equilibria require unrealistic coordination among symmetric players.

Figure 3: Under-report rate over 30 periods



the audit frequency turns out to be significantly higher ($p < 0.05$), due to the fact that there are fewer "low-income" reports needed to be audited. The audit success rate is remarkably higher as well ($p < 0.05$), leading the lying H-type taxpayers a significantly lower payoff than the *Traditional* treatment ($p < 0.05$). The average budget usage ratio is 95.63%, which is again significantly higher than that under the traditional rule (32.03%).

Result 4 *The non-compliance rate in the Strong-Economy treatment is significantly lower than both the Traditional and the Bounded treatments. This high deterrence rate is acheived with a significantly lower implementation cost, and a high budget usage ratio.*

In the last subsection, we showed that behavior in both treatments is consistent with a loss aversion model with stochastic decision errors. To examine how they explain pattern in this treatment, we do the following exercise. We first estimate the perceived audit probability of the *Strong-Economy* treatment given the α, β and λ parameters obtained in the *Bounded* treatment. Then we use all thse information to calculate the predicted under-reporting rate of the treatment $\widehat{b}^{SE}$. It turns out that $\widehat{b}^{SE}$ is 33.80%, which is again very close to the actual prediction 33.95%. This is an indication that behavior in the *Strong-Economy* treatment is again consistent with the loss-aversion model with decision errors.

Interestingly, the preceived probability in the *Strong-Economy* treatment, 0.344, is only mildly larger than that in the *Bounded* treatment, 0.305. That means given a five percentage increase in audit probability perception leads to a 23 percent increase in compliance level. This asymmetry lies in the fact that subjects value gains and losses differently to the reference point. When the perceived audit probability is 0.305 in the *Bounded* treatment, the value of under-reporting is in the gain domain, and marginally larger than 0, the value of the reference point. When this probability increases to 0.344 however, the value of under-reporting falls into the loss domain with a larger distance to the reference point. Since subjects are loss-averse, they weight losses more than gains, and hence lower under-reporting frequency more drastically.

How does loss aversion link to coordination failure? If taxpayers are expected profit maximizers, they will underreport as long as they think the probability that others are going to underreport is larger than 0.432. It would be much harder, on the other hand, for loss-averse players to choose to underreport. Given that they are more sensitive to losses than gains, they will choose to underreport only when they think the other H-types are going to under-report with probability higher than 0.774. This threshold requires more coordination among taxpayers, and involves a substantially higher degree of strategic uncertainty.

According to Brandenburger (1996) definition, strategic uncertainty arises when there is "uncertainty concerning the purposeful behaviour of players in an interactive decision situation", as opposed to a game against nature. Strategic uncertainty is widely documented in the many experimental studies such as the coordination games (e.g. Huyck et al. (1990), Huyck et al. (1991)) and market entry games (e.g. Sundali et al. (1995); Erev and Rapoport (1998)). Recently, Heinemann et al. (2009) propose a method to measure strategic uncertainty by eliciting certainty equivalents analogous to measure risk attitudes in lotteries. In their experiment, N subjects have to choose simultaneously between a series of lotteries pairs. In each pair, lottery A always yields a sure flat payoff, and lottery B yields a payoff if the minimum number of players selected is k. They find that the number of B-choices in coordination games decreases with an increasing coordination requirement k. If holding k constant, N has a strong positive effect on coordination, since a large N reduces the relative hurdle to coordination. These behavioral patterns indicate that subjects are strategic uncertainty averse. Applying the study by Heinemann et al. (2009), we argue that the risk-dominant equilibrium is more likely to be chosen by the loss-averse subjects than expected profit maximizers.

5.4 Learning and Social demographics

In the post-questionnaire, we ask subjects their social background information such as gender and nationality. This information allows us to study how subjects form and adjust their under-reporting decisions under different rules. The first specification concerns compliance behavior. We use the following random-effect probit model specification:

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}$$

where y equals 1 if subjects decide to under-report, and equal to 0 otherwise. Furthermore, x is a vector of explanatory variables, the α_i represent individual random effects and β is a vector of parameters. The explanatory variables include subjects' social backgrounds such as gender, nationality and experience of economics, under-report performance in last period, time and its square term.

Besides compliance behavior, we also investigate how individual characteristics and previous performance influence players' perceived audit probability of the bounded rules. To do that, we re-run loss aversion model in section 5.2 and allow the perceived a parameter in model 2 depends social background information: $\widehat{a} = \beta x + \varepsilon_{it}$. The results of the two specifications are shown in Table 3.

Table 3: Under-report influences of social background and learning

	Compliance behavior			Audit Probability Perception	
	Traditional	*Bounded*	*Strong-Economy*	*Bounded*	*Strong-Economy*
Under-report Detection Experience	-0.102	-0.498***	0.010	0.075**	0.030***
	(0.186)	(0.195)	(0.105)	(0.038)	(0.006)
Period	-0.039	-0.040	-0.082***	0.008	0.002***
	(0.026)	(0.032)	(0.023)	(0.007)	(0.006)
Period^2	0.0006	0.0004	0.001	-0.0001	-0.0002
	(0.0008)	(0.0009)	(0.0006)	(0.0002)	(0.0002)
Gender (1 for men)	0.756*	0.960*	0.459	-0.020	0.003
	(0.413)	(0.544)	(0.478)	(0.017)	(0.011)
Years of learning economics	0.052	0.797***	0.278	-0.015	-0.001
	(0.186)	(0.287)	(0.214)	(0.014)	(0.005)
Econ experience × Game-theory	0.083	-0.574**	-0.066	0.012	-0.0009
	(0.144)	(0.274)	(0.187)	(0.013)	(0.004)
Dummy for Eastern Europeans	0.015	-0.409	0.125	0.043**	-0.020
	(0.899)	(0.922)	(0.925)	(0.022)	(0.201)
Dummy for Dutch	-0.102	1.190	0.620	0.004	-0.025
	(0.691)	(0.777)	(0.677)	(0.019)	(0.018)
Dummy for Chinese	-0.234	1.336	-0.009	-0.003	-0.014
	(0.652)	(0.798)	(0.778)	(0.019)	(0.023)
Dummy for other Asian	-0.878	0.159	0.940	0.022	-0.026
	(0.784)	(0.971)	(0.968)	(0.025)	(0.023)
Mach-IV Scale	0.025	-0.019	0.018	0.003	-0.0003
	(0.016)	(0.018)	(0.015)	(0.003)	(0.004)
Tax Filing Experience	-0.005	-0.304	-0.339	0.011	0.001
	(0.458)	(0.492)	(0.575)	(0.012)	(0.011)
Constant	-0.508	5.318**	0.918	0.209***	0.305***
	(2.091)	(2.308)	(1.800)	(0.042)	(0.038)
Log-likelihood	-444.868	-359.410	-640.430	-997.726	-1437.232
Observations	957	912	1670	2256	3014

Notes: *10% significance; **5% significance, ***1% significance. We only include observations which players receive high-income. Standard errors are clustered on individuals.

It is clear that learning affect subjects in different ways across treatments. In the *Bounded* treatment, detection experience in the previous round decreases non-compliance propensity. Interestingly, players with economics background are more likely to under-report, an evidence suggesting that training in economics might results in behavior more in line with homo economicus. These effects, however, do not exist in the other two treatments. In both *Traditional* and *Bounded* treatments, men are more likely to under-report than women. However, this is not the case in the *Strong-Economy* treatment. In fact, no other social demographic information affects behavior except for time. When we let perceived audit probability depends on social demographic information under the bounded rule treatments, the only variable significant is the lag audited experience. That is, if a cheater was caught in the previous period, the perceived audit probability increases. This pattern is largely in line with reinforcement learning. The invariant influences for social background information might indicate that the effect of bounded rules are robust and does not subject to much change with respect to subjects' social background information, especially in the *Strong-Economy* treatment.

6 Conclusion

In this study, we compare the deterrence effect of a new auditing rule, the bounded rule, to the traditional rule in a controlled laboratory environment. In a tax compliance game, subjects are endowed with a high- or low-income with a certain probability. On being notified a pre-specified audit rule (traditional or bounded), their task is to report the type of income to the auditor. Individual profit maximization and non-cooperative game theory suggest that the two rules induce the same level of compliance. We find that the compliance rate in the bounded rule is indeed the same as that in the traditional rule. Given the same compliance level induced, the bounded rule is more cost-effective in terms of both implementation cost and budget usage ratio.

Although the theory predicts correctly the directions regarding treatment difference, it does not do well in predicting the absolute levels. We find that compliance rates are higher than the prediction in both treatments, even taking into account the fraction of intrinsically honest players. In both treatments, at least half of the subjects comply to a certain extent, but switch their choices alternatively. To explain behavioral anomalies, we introduce discrete choice models, within which we compare several alternative models. We find that loss aversion combining with stochastic errors are more successful at tracking observed data patterns.

In a bounded rule with multiple equilibria, the subjects fail to coordinate on the payoff-dominant outcome. In both bounded rule treatments, the total number of audits is significantly lower than in the traditional rule treatment. Moreover, the bounded rule commits fewer resources (2 audits versus 8 audits for the traditional rule), and uses these commited resources more efficiently. Hence, we can safely conclude that the bounded rule is a more cost-effective audit mechanism compared to the traditional rule.

What can we learn from this experiment? The experiment data strongly favors the bounded rule to be a superior audit selection rule. The first advantage is that it helps the auditors to plan budget more efficiently and more precisely. Many large organizations such as National Aeronautics and Space Administration (NASA), US Internal Revenue Service (IRS) and Atomic Energy Agency

(IAEA) typically requires divisions to have their budget plans approved before they can spend money on activities in the following year. A more efficient planning like the bounded rule can pin down the exact budget needed for the activity of auditing, and hence decreases the unnecessarily reserved resources. The second advantage is that the bounded rule is cheaper to use. When the percentage of "red-flag" reports are large, the bounded rule needs only fewer implementation resources in achieving the same level of compliance. It might be useful in combating self-employed firms in developing countries where the government has a constrainted budget to audit the unpaid taxes. Moreover, the bounded rule performs even better when the income distribution is right-skewed. Combined with the fact that the bounded rule requires fewer resources, it could be particularly useful for developed countries with high cost of employing audit manpower.

The use of experiment also helps studying behavior under the two selection rules. We find that people act like loss-averse with stochastic errors in such settings, rather than expected profit maximizers. History of play also affects their perception toward audit probability under the bounded rule. Incorporating this evidence would better help tax administration to adjust their policies to encourage people in paying their taxes in a more cost-effective way.

This is just a first step into the investigation of the bounded rule. In our current setup, taxpayers can only decide whether to under-report or honestly report. In our future study, we could extend the model so that they can also specify how much to under-report. Another possible extension is to introduce human auditor to further examine the strategic interactions. Taxpayers are able to communicate with each other in reality. Alm and McKee (2004) have shown that such cheap-talk communication could help taxpayers to coordinate on non-compliance (payoff-dominant) outcome. However, if a strategic auditor could observe this, he is able to adjust the audit capacity accordingly to combat collusion among taxpayers.

References

Alm, J., M. B. Cronshaw, and M. McKee (1993). Tax compliance with endogenous audit selection rules. *Kyklos 46*(1), 27–45.

Alm, J., B. R. Jackson, and M. McKee (2009, April). Getting the word out: Enforcement information dissemination and compliance behavior. *Journal of Public Economics 93*(3-4), 392–402.

Alm, J. and M. McKee (1998, August). Extending the lessons of laboratory experiments on tax compliance to managerial and decision economics. *Managerial and Decision Economics 19*(4/5), 259–275.

Alm, J. and M. McKee (2004, July). Tax compliance as a coordination game. *Journal of Economic Behavior & Organization 54*(3), 297–312.

Bayer, R. and F. Cowell (2009). Tax compliance and firms' strategic interdependence. *Journal of Public Economics Volume 93, Issues 11-12*, 1131–1143.

Blackwell, C. (2007). A meta-analysis of tax compliance experiments. In Martinez-Vazquez and J. Alm (Eds.), *Tax Compliance and Evasion, forthcoming.*

Boylan, S. J. and G. B. Sprinkle (2001, March). Experimental evidence on the relation between tax rates and compliance: The effect of earned vs. endowed income. *Journal of the American Taxation Association 23*(1), 75–90.

Brandenburger, A. (1996, May). Strategic and structural uncertainty in games. In R. J. Zeckhauser, R. L. Keeney, and J. K. Sebenius (Eds.), *Wise Choices: Decisions, Games, and Negotiations*, pp. 221–232. Harvard Business Press.

Chen, R. and R. M. Johnston (2008). Securities and exchange commission comment letters: Enforcing accounting quality and disclosure. *SSRN working paper*.

Christie, R. and F. Geis (1970). *Studies in Machiavellianism.* Academic Press, New York.

Dhami, S. and A. al Nowaihi (2007, September). Why do people pay taxes? prospect theory versus expected utility theory. *Journal of Economic Behavior & Organization 64*(1), 171–192.

Dhami, S. and A. Al-Nowaihi (2010, August). Optimal taxation in the presence of tax evasion: Expected utility versus prospect theory. *Journal of Economic Behavior & Organization 75*(2), 313–337.

Elffers, H. and D. J. Hessing (1997, April). Influencing the prospects of tax evasion. *Journal of Economic Psychology 18*(2-3), 289–304.

Erev, I. and A. Rapoport (1998). Coordination, 'magic,' and reinforcement learning in a market entry game. *Games and Economic Behavior 23*(2), 146–175.

Fischbacher, U. (2007). Z-tree - zurich toolbox for readymade economic experiments. *Experimental Economics 10(2)*, 171–178.

Fischbacher, U. and U. Stefani (2007, May). Strategic errors and audit quality: An experimental investigation. *The Accounting Review 82*(3), 679.

Graetz, M., J. Reinganum, and L. Wilde (1986). The tax compliance game: Toward an interactive theory of tax enforcement. *J. Law Econ. Organ. 2*, 1–32.

Heinemann, F., R. Nagel, and P. Ockenfels (2009). Measuring strategic uncertainty in coordination games. *Review of Economic Studies 76*(1), 181–221.

Holt, C. A. and S. K. Laury (2002, December). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Huyck, J. B. V., R. C. Battalio, and R. O. Beil (Aug., 1991). Strategic uncertainty, equilibrium selection, and coordination failure in average opinion games. *The Quarterly Journal of Economics 106*(3), 885–910.

Huyck, J. B. V., R. C. Battalio, and R. O. Beil (Mar., 1990). Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review 80*(1), 234–248.

Kahneman, D. and A. Tversky (1979, March). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–91.

Kim, C. K., J. H. Evans, and D. V. Moser (2005, October). Economic and equity effects on tax reporting decisions. *Accounting, Organizations and Society 30*(7-8), 609–625.

Kim, C. K. and W. S. Waller (2005). A behavioral accounting study of strategic interaction in a tax compliance game. pp. 35–56.

King, S. and S. Sheffrin (2002, August). Tax evasion and equity theory: An investigative approach. *International Tax and Public Finance 9*(4), 505–521.

Lennox, C. and J. Pittman (2010). Auditing the auditors: Evidence on the recent reforms to the external monitoring of audit firms. *Journal of Accounting and Economics 49*(1-2), 84 – 103.

Loomes, G. (2005, December). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics 8*(4), 301.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers of Econometrics New York: Academic press.*

McFadden, D. (2001, June). Economic choices. *The American Economic Review 91*(3), 351–378.

McKelvey, R. D. and T. R. Palfrey (1995, July). Quantal response equilibria for normal form games. *Games and Economic Behavior 10*(1), 6–38.

Moser, D. V., J. H. Evans, and C. K. Kim (1995, October). The effects of horizontal and exchange inequity on tax reporting decisions. *Accounting Review 70*(4), 619–634.

Nelson, M. and H.-T. Tan (2005). Judgment and decision making research in auditing: A task, person, and interpersonal interaction perspective. *Auditing: A Journal of Practice & Theory 24*, 41–71.

Reinganum, J. F. and L. L. Wilde (1985, February). Income tax compliance in a principal-agent framework. *Journal of Public Economics 26*(1), 1–18.

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1446–1465.

Slemrod, J. and S. Yitzhaki (2002). Tax avoidance, evasion, and administration. *3*, 1423–1470.

Starmer, C. and R. Sugden (1991). Does the random-lottery incentive system elicit true preferences? an experimental investigation. *The American Economic Review 81*(4), 971–978.

Sundali, J. A., A. Rapoport, and D. A. Seale (1995). Coordination in market entry games with symmetric players. *Organizational Behavior and Human Decision Processes 64*(2), 203 – 218.

Tversky, A. and D. Kahneman (1992, October). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

US Department of the Treasury, U. A. (2006). Fy 2007 budget in brief. *Washington, DC.*

Wilcox, N. T. (2009). "stochastically more risk averse": A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics forthcoming*.

Yaniv, G. (1999, December). Tax compliance and advance tax payments: a prospect theory analysis. *National Tax Journal 52*(4), 753–764.

Yim, A. (2009, December). Efficient committed budget for implementing target audit probability for many inspectees. *Management Science 55*(12), 2000–2018.

Zimbelman, M. F. and W. S. Waller (1999). An experimental investigation of Auditor-Auditee interaction under ambiguity. *Journal of Accounting Research 37*, 135–155.

Appendix

A Proofs

A.1 Proof of Proposition 1

The proof is trivial that reporting high-income is a dominated strategy for the L-type players. To prove that H-type players have a undominated strategy of under-reporting given that L-type players comply dominance, we need to show that the expected payoff from underreporting will be strictly larger than the sure payoff from reporting truthfully. Moreoever, this holds regardless of the beliefs H-type players $b_i \in [0,1]$ towards the other H-type players. For simplicity, we assume that every player is homogeneous so that $b_i = b$. The sure payoff of reporting truthfully is 12.5. The expected payoff from underreporting is:

$$
\begin{aligned}
E(\pi_{lie}) &= \sum_{n=0}^{N-1} \mathbf{Bin}(n; N-1, B_i) \times \{\min(\frac{2}{n+1}, 1) \times \pi_F + [1 - \min(\frac{2}{n+1}, 1)] \times \pi_S\} \\
&= \pi_S - (\pi_S - \pi_F) \times \sum_{n=0}^{N-1} \mathbf{Bin}(n; N-1, B_i) \times \min(\frac{2}{n+1}, 1) \\
&= 22.5 - 20 \times \sum_{n=0}^{N-1} \mathbf{Bin}(n; N-1, B_i) \times \min(\frac{2}{n+1}, 1)
\end{aligned}
$$

Where $B = (1-q)+qb$ given previous discussion. Since $b \in [0,1]$ and $q = 0.5$, $B \in [\frac{1}{2}, 1]$. In order to show that $E(\pi_{lie})$ is strictly larger than 12.5, we need to prove $\sum_{n=0}^{N-1} \mathbf{Bin}\ (n; N-1, B_i) \times \min(\frac{2}{n+1}, 1)$ is smaller than 0.5 for any $B \in [\frac{1}{2}, 1]$.

Rearranging $Temp = \sum_{n=0}^{N-1} \mathbf{Bin}\ (n, N-1, B_i) \times \min(\frac{2}{n+1}, 1)$, we have:

$$
\begin{aligned}
Temp &= \mathbf{Bin}(0,7,B) + \mathbf{Bin}(1,7,B) + \sum_{n=2}^{N-1} \mathbf{Bin}(n, N-1, B) \times (\frac{2}{n+1}) \\
&= \mathbf{Bin}(0,7,B) + \mathbf{Bin}(1,7,B) + \sum_{n=2}^{N-1} \mathbf{Bin}(n, N-1, B) \times (1 - \frac{n-1}{n+1}) \\
&= \sum_{n=0}^{N-1} \mathbf{Bin}(n, N-1, B) - \sum_{n=2}^{N-1} \mathbf{Bin}(n, N-1, B) \times (\frac{n-1}{n+1}) \\
&= 1 - \sum_{n=2}^{N-1} \mathbf{Bin}(n, N-1, B) \times (\frac{n-1}{n+1})
\end{aligned}
$$

Therefore, we need to show that $\sum_{n=2}^{N-1} \mathbf{Bin}\ (n, N-1, B) \times (\frac{n-1}{n+1})$ is larger than 0.5 for any given

$B \in [\frac{1}{2}, 1]$. By plugging $q = 0.5$ and $N = 7$, spreading and re-arranging the each term in the binomial distribution, we have the following polynomial with degree seven: $F = 7B^2 - 17.5B^3 + 21B^4 - 14B^5 + 5B^6 - 0.75B^7$. Taking the first derivative yields a polynomial with degree six:

$$\begin{aligned}\frac{\partial F}{\partial B} &= -5.25B^6 + 30B^5 - 70B^4 + 84B^3 - 52.5B^2 + 14B \\ &= (5.25B^5 - 5.25B^6) + (24.75B^5 - 70B^4 + 49.5B^3) + (34.5B^2 - 52.5B^2 + 14B)\end{aligned}$$

The first term, $(5.25B^5 - 5.25B^6)$, is always non-negative if $B \in [\frac{1}{2}, 1]$. It is not difficult to show that the second and the third terms are always positive for any given $B \in [\frac{1}{2}, 1]$. Hence, $\frac{\partial F}{\partial B} > 0$, which means that $\sum_{n=2}^{N-1} \textbf{Bin}\ (n; N-1, B) \times (\frac{n-1}{n+1})$ is increasing in B for any $B \in [\frac{1}{2}, 1]$. When $B = 0.5$, $F_{\min} = 0.5098$, which is strictly larger than 0.5. Hence the H-type players choose to underreport conditional on the dominant strategy of the L-type players. ∎

Next, we prove that the introduction of intrinsically honest players does not change the directions of treatment difference. Let ρ be the proportion of honest players, and $1-\rho$ be the proportion of strategic, self-interest profit maximizers, $0 < \rho < 1$. We assume that the ρ is the same in both treatments. Now we only need to show that the existence of intrinsically honest players does not affect the strategy of profit maximizers. When the strategic players are assigned to be L-type, they gain a higher payoff by reporting truthfully, regardless of the auditing rule implemented. In the *Traditional* treatment, H-type profit maximizers only compares a sure payoff of reporting truthfully and the tax evasion gamble if they underreport. Hence, the existence of honest players will not affect their choices. In the *Bounded* treatment, the subjective beliefs of strategic, H-type players of the number of "low-income" reports now become: $B = (1-q) + q(1-\rho)b$. Given that $q = 0.5$, $0 < \rho < 1$, B still lies in the interval $[\frac{1}{2}, 1]$. Therefore, the remaining analysis follows the proof of Proposition 1 above.

In the presence of honest players, the non-compliance rate of both treatments will be$1-\rho$.

A.2 Equilibrium analysis for *Strong-Economy* Treatment

Let $\sigma_i(j)$ be the probability type i player (H-type or L-type) will strategy j (u or h). There are two pure Nash equilibria and one mixed-strategy equilibrium in this treatment:

$$\{(\sigma_H(u) = 1, \sigma_L(h) = 1), (\sigma_H(h) = 1, \sigma_L(h) = 1), (\sigma_H(u) = 0.432, \sigma_L(h) = 1)\}$$

In words, the two pure NEs are 1) all H-type players under-report and 2) all H-type players honestly report. L-type players always honestly report.

Let us examine the former case. Given a H-type player thinks that all other H-type players choose strategy u, he will have an expected payoff of 17.5 by playing strategy l. By deviating to h, the payoff decreases to 12.5. Since we assume symmetry among players, no one has the incentive to deviate from under-reporting, which constitutes a NE. A highly similar analysis applies to the latter case. Given that all other H-type players play strategy h, a strategy deviation from h to l will yield a lower expected payoff for H-type players (from 12.5 to 3.59). Hence no one has the incentive to deviate.

On top of the two pure equilibria, there is also a mixed-strategy equilibrium in which each H-type taxpayer is indifferent between honest-report and under-report. Given that the game parameters, the under-report probability b which induces utility indifference is $b^*_{SE} = 0.432$. ∎

Note that there are other asymmetric equilibria in the game. However, we ignore them in a symmetric setting, since these equilibria require unrealistic coordination among symmetric players.

B Instructions

B.1 Instructions Comparison

The instructions given in the next subsection are for the *Bounded* treatment. It differs from the instructions for the other treatments as follows:

- *Traditional* treatment

  1. The second bullet (concerning matching protocol) of the list under "Task Description" in the instructions for the "Tax Compliance Game" is absent.
  2. The "Audit Probability Table" is absent.
  3. The phrase "see audit prob. table" in the "Payoff Table" becomes 0.4.

- *Strong-Economy* treatment

  1. In the third bullet of the list under "Task Description" in the instructions for the "Tax Compliance Game", the probability of receiving 25 becomes 0.9, and accordingly the probability of receiving 10 becomes 0.1.
  2. In the "Payoff Table" (immediately before "Payment Method" in the instructions for the "Tax Compliance Game"), the probabilities in the second column become 0.9 and 0.1, respectively.

B.2 Instructions for *Bounded* Treatment

- Please read these instructions carefully!
- Please do not talk to your neighbours and remain quiet during the entire experiment.
- If you have a question, please raise your hand. We will come to you to answer it.
- You will receive a show-up fee of €3 for completing all tasks in the experiment, independent of your performance.

Task Description

- This session consists of 30 periods of play; each period is completely independent of the others.

- Of the participants in the room, two groups of 8 participants will be randomly formed at the beginning of each period. You will not know the identity of the other players in your group in any period.
- At the beginning of each period, you will receive a taxable income of either €25 or €10. The probability of receiving €25 is 0.5; and the probability of receiving €10 is 0.5.
- Your task is to report your income to the auditor, which is played by a computer. The amount that you report is your decision. You can report either €25 or €10, regardless of your received income.

After-tax Income Determination

Your after-tax income in this period is determined by the following two steps: tax payment and an audit.

Step One: Tax payment

The tax rate is 50% for those who reported €25 and 25% for those who reported €10. Suppose the income you received is €25:

- If you report €25 to the auditor, the auditor will charge €12.5 (50% of €25) as tax. So your after-tax income in this period equals to €25 – €12.5 = €12.5.
- If you report €10 to the auditor, the auditor will charge €2.5 (25% of €10) as tax. So your after-tax income in this period equals to €25 – €2.5 = €22.5.

Suppose the income you received is 10:

- If you report €10 to the auditor, the auditor will charge €2.5 (25% of €10) as tax. So your after-tax income in this period equals to €10 – €2.5 = €7.5.
- If you report €25 to the auditor, the auditor will charge €12.5 (50% of €25) as tax. So your after-tax income in this period equals to €10 – €12.5 = -€2.5.
- In sum, the auditor charges tax based on your reported income, instead of your received income.

Step Two: Audit

The auditor does not know your received income unless your report is later audited.

Auditing procedure:

- If your reported income is €25, it will not be audited. That means what you have earned in step one (€12.5 or -€2.5) will be your after-tax income (if received income is 25 and 10 respectively).

- Regardless of your received income, if your reported income is €10, there is a chance that your report will be audited. The outcome is as follows:

- Suppose your reported income is €10 AND your received income is also €10. Then what you have earned in step one (€7.5) will be your after-tax income no matter whether your report is audited or not.

- Suppose your reported income is €10 AND your received income is €25. If your report is not audited, you will keep the €22.5 earned in step one; if audited, you will get €2.5.

Auditing probability:

The number of reports the auditor will audit depends on the number of players reporting an income of €10 in a group.

- If the number of €10 income report is equal to two or less, the auditor will audit all €10 reports.

- If the number of €10 income report is three or more, then two out of such reports will be selected for audit randomly.

- The following table named "Audit Probability Table" shows the audit probabilities for a player who reported an income of €10.

Audit Probability Table

Number of €10 reports	0	1	2	3	4	5	6	7	8
Audit Probability	100%	100%	100%	66.7%	50%	40%	33.3%	28.6%	25%

- The following table named "Payoff Table" summarizes all possible scenarios you may encounter in one period and the related payoffs:

Payoff Table

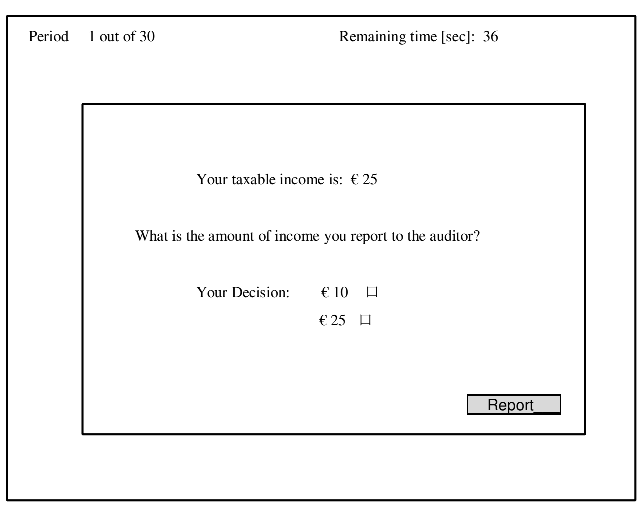
Received Income	Probability	Reported Income	Audit Probability	After-tax Income if audited	After-tax Income if NOT audited
€25	0.5	€25	0	€12.5	€12.5
		€10	see audit prob. table	€2.5	€22.5
€10	0.5	€10	see audit prob. table	€7.5	€7.5
		€25	0	$-$€2.5	$-$€2.5

Payment Method

- At the end of this experiment, one out of 30 periods will be selected to determine your payoff for this task. The computer program will generate a random number from 1 to 30. This number will determine one of the 30 periods. Your performance in that period determines your payoff.
- You will be paid in based on your after-tax income for the randomly selected period.
- Because each period is equally likely to be selected for payment determination, you should make your decision in each period as if that period would be selected for payment.
- Your payoff will be paid out in cash at the end of the experiment along with your earnings in the other task(s).

We will now explain how the computer screens look like.

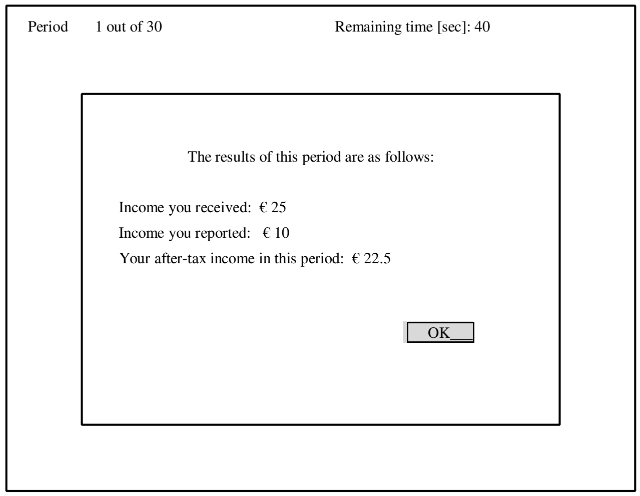
SCREEN 1



Here you can decide the amount of income to report to the auditor. Please select either "€10" or "€25", and confirm your choice by pressing the "Report" button.

Warning: Before pressing the button, make sure your choice is correct. You cannot change your decision after you have pressed OK.

SCREEN 2



This is the feedback table you will receive regarding your after-tax income. Your will find information on the initial taxable income you received, the income you reported and your after-tax income in this period.

Click on OK when you finish checking the information.

Note that the purpose of the screen shots is to clarify the procedure, rather than provide advice about how to act. You should make the decisions that are best for you.

Please raise your hand if you have any questions at this moment.

B.2.1 Risk Elicitation Task[15]

Task Description

In this task, you are asked to make decisions to 21 choice pairs. In each choice pair, you need to select between two lotteries labeled "Lottery A" and "Lottery B". Please, take your time and read each choice pair carefully. An example of a typical choice pair is given below:

Choice No.1	Lottery A	€5.5 with probability 0.5 or €3.5 with probability 0.5	Your choice:	Lottery A ☐
	Lottery B	€9 with probability 0.5 or €0.5 with probability 0.5		Lottery B ☐

Payment Method

- You need to make choices for all 21 choice pairs. However, only one of the 21 choices you have made will be chosen for the payoff determination of this task. First, the computer program will generate a random number from 1 to 21. This number will determine a choice pair. Then, the computer program will simulate the lottery you have chosen and reveal the outcome on your screen. The outcome of this lottery will determine your payoff.

- For example, suppose that the computer program has generated a random number 2. It will then check what you have selected in choice pair number 2. Suppose that you have chosen Lottery A in that choice pair. Then the computer program will simulate Lottery A and reveal your payoff (either €5.5 or €3.5). Your payoff will be paid out in cash at the end of the experiment along with your earnings for the other task.

It is important that you fully understand the lottery selection task. Please raise your hand if you have any questions at this moment.

[15] The risk elicitation task is conducted after the tax compliance game. However, the subjects do not know the existence of this task when they were playing the tax compliance game.

B.2.2 Post-experimental Questions

Questions on Treatment Manipulation

Please evaluate the following statements with respect to the tax reporting task:[16]

1=strongly disagree, 2=somewhat disagree, 3=slightly disagree, 4=no opinion, 5=slightly agree, 6=somewhat agree, 7=strongly disagree

1. The instructions were clearly formulated.
2. I felt that I performed well on the task.
3. I received plenty of time to carry out the task.
4. I was motivated to do well on the task.
5. The task was fun to perform, motivating me to achieve a payoff as high as possible.
6. I considered the tax reporting task as fairly complex.
7. My payoff is determined not only by my own decision, but decisions of the other players.
8. When making my decision, I thought about what other players might do.
9. I feel obliged to report the received income in each period.
10. The chance I have received €25 is about 50%.[17]

Questions on Background Information

Please answer the following survey questions. Your answers will be used for this study only. Individual data will not be exposed.

1. What is your gender?
2. What is your nationality?
3. How many years have you already studied in economics?
4. Have you have learnt a course related to game theory?
5. Have you ever had a part-time job?

[16] The first five questions are used to understand the subjects' perception about the experimental setup and instructions in general. We do not expect to find differences across treatments. The last five questions focus on capturing different types of manipulations of the treatments; therefore, we expect to see differences across manipulations.

[17] In the *Strong-Economy* treatment, the chance should be 90%, instead of 50%.

Quesitons on Mach IV Scale[18]

In the following you will find a list of statements. Please read them carefully and answer them to what extent you agree or disagree. Even if in some cases you would like to say, that depending on the circumstances, you should only choose one of the answers. Since all responses are anonymous you can indicate without any worry. There is nobody on who you need to make a good impression. Only if you answer very honestly can the results be used.

> *1=strongly disagree, 2=somewhat disagree, 3=slightly disagree, 4=no opinion, 5=slightly agree, 6=somewhat agree, 7=strongly disagree*

1. Never tell anyone the real reason you did something unless it is useful to do so.
2. The best way to handle people is to tell them what they want to hear.
3. One should take action only when sure it is morally right.
4. Most people are basically good and kind.
5. It is safest to assume that all people have a vicious streak and it will come out when they are given a chance.
6. Honesty is the best policy in all cases.
7. There is no excuse for lying to someone else.
8. Generally speaking, people won't work hard unless they're forced to do so.
9. All in all, it is better to be humble and honest than to be important and dishonest.
10. When you ask someone to do something for you, it is best to give the real reasons for wanting it rather than giving reasons which carry more weight.
11. Most people who get ahead in the world lead clean, moral lives.
12. Anyone who completely trusts anyone else is asking for trouble.
13. The biggest difference between most criminals and other people is that the criminals are stupid enough to get caught.

[18] Question 3, 4, 6, 7, 9, 10, 11, 14, 16 and 17 are reverse coded.

14. Most people are brave.
15. It is wise to flatter important people.
16. It is possible to be good in all respects.
17. Barnum was wrong when he said that there's a sucker born every minute.
18. It is hard to get ahead without cutting corners here and there.
19. People suffering from incurable diseases should have the choice of being put painlessly to death.
20. Most people forget more easily the death of their parents than the loss of their property.