

Regulatory Performance of Audit Tournaments and Compliance Observability

Timothy N. Cason
Purdue University
cason@purdue.edu

Lana Friesen
University of Queensland
l.friesen@uq.edu.au

Lata Gangadharan
Monash University
Lata.Gangadharan@monash.edu

December 2014

Abstract

This paper reports an experimental study of two stochastic audit schemes for enforcing regulatory compliance. In the Random Audit mechanism firms are randomly chosen for inspection. In the Tournament Audit mechanism the probability of inspection increases with the degree of estimated under-reporting. The experiment also varies the observability of identity, output, and compliance decisions. Optimal output is theoretically independent of the auditing schemes, but equilibrium reporting is higher under the Tournament mechanism than Random auditing. Experimental findings are consistent with the theoretical predictions for reporting, but not for output. In particular, we find that average output is lower and reporting is higher in the Tournament treatment compared to the Random Audit treatment. At the individual level, a majority of participants misreported in most periods. Social observability does not affect output or reporting significantly in either of the audit treatments.

Keywords: Auditing, Rank-Order Tournament, Laboratory Experiment, Social Observability, Regulatory Enforcement, Tax Compliance.

JEL Classification: H41, L51, Q58

1. Introduction

Auditing is important for ensuring that the regulatory programs instituted by governments are effective and that individuals and organizations are appropriately compliant with regulations. Auditing however is expensive and consumes scarce regulatory budgets. Policy-makers can therefore benefit from audit mechanisms that improve compliance cost-effectively. This paper presents one such mechanism that focuses on competition, in which firms compete with each other to avoid being audited by the regulator. An individual firm's probability of being audited depends on its own actions as well as the actions of the other similar firms in the industry or region. Our primary research objective is to examine if this endogenous tournament mechanism leads to higher compliance rates than a random audit mechanism.

Firms and individuals may also be motivated by their public image. For example, they may suffer disutility when being revealed as violators of social norms (Andreoni and Petrie, 2004; Coricelli et al., 2010) or adverse market valuations when poor compliance is publicly revealed (Konar and Cohen, 2001). Therefore, making actions public is another potential method for improving compliance with regulatory goals. If individuals or firms are motivated by shame, prestige, or status in social settings their desire for social approval could lead to more compliance. Reputational concerns and customer, employee and shareholder preferences for investments in corporate social responsibility may influence compliance behavior (Kitzmueller and Shimshack, 2012). Effectively communicating compliance with different regulatory goals, such as honest reporting of taxable income, eschewing creative (or fraudulent) accounting practices to shelter income, and complying with product, workplace and environmental safety measures can affect corporate image. For policy makers, the positive impact of social observability on compliance rates could create significant benefits, especially if the costs

associated with this are mostly negligible. The secondary goal of this paper is therefore to investigate whether increasing the visibility of actions can lead to improved compliance.

We conduct an experiment to evaluate the effectiveness of audit tournaments and social observability as compliance mechanisms. Laboratory experiments are a useful methodology for investigating the effects of alternative policy options. For mechanisms that are innovative and not yet implemented in the field, it is difficult to find effective counterfactuals, making empirical evaluation impossible. In addition to comparing performance across different schemes, comparing the experimental outcomes with equilibrium predictions in a controlled environment further allows us to evaluate empirically the insights arising from the theory. The experimental approach also allows for random assignment of treatment conditions, which permit unambiguous inferences regarding causality.

We consider a scenario in which firms choose an output level, with higher choices providing private benefits but imposing negative externalities on other members of society. These negative externalities lead to direct, negative payoff consequences due to others' increased output, and they are a primary reason why public exposure may motivate more socially-beneficial choices. Firms are required to report their output to the regulator but reporting incurs a cost, for example through a per-unit tax. To avoid this reporting cost, firms may choose to misreport and enjoy private benefits from a high output. Output is not easily observable and to better align aggregate output to what is socially optimal by accounting for the externalities, regulators need to audit or inspect the firms. This scenario applies in several areas of the economy where regulatory effort is exerted by governments to achieve social outcomes, such as environmental and health and safety regulations, as well as sales, VAT and income tax reporting.

We consider two auditing schemes, varied exogenously across experimental treatments. In the Random Audit treatment firms are randomly inspected with a constant exogenously determined probability. In the Tournament Audit treatment, firms are ranked based on their reporting. The probability of inspection in this treatment increases with the estimated degree of under-reporting. Hence firms have an incentive to compete with each other in terms of compliance, since lower under-reporting compared to others in the industry or regulatory group results in lower chances of an audit. This competition amongst firms can be characterized as a Rank-Order Tournament (Lazear and Rosen, 1981). Firms that are found to be non-compliant incur identical fines in both treatments.

To examine the influence of social visibility and observability we implement the two audit treatments described above with two levels of information disclosure. In the Low information condition participants only learn the reporting choices of others, who always remain anonymous. In the High information scenario we add feedback about actual output and compliance of the inspected group members, and also display digital photographs of all the participants. Theoretically, the social visibility of actions and identities should not influence output and reporting choices in any of the treatments, however based on previous research (discussed below) we anticipate that making decisions public and displaying photos could potentially lead to participants internalizing the externalities they cause in their decision making. Further the impact of visibility could differ across audit treatments.

Results from our stylized theoretical model show that the optimal level of output is independent of the auditing scheme chosen by the regulator. Equilibrium reporting, however, is higher under the Tournament mechanism than the Random Audit scheme. Findings from our experiment are consistent with the theoretical predictions for reporting, but not for output. In

particular, we find that average output is lower and reporting is higher in the Tournament treatment as compared to the Random Audit treatment. At the individual level, a majority of subjects misreported in most periods. Social observability does not affect output or reporting significantly in either of the audit treatments.

Our paper contributes to two strands of literature. The first is a more established literature on monitoring and auditing. Within this literature, our work is most closely related to Gilpatric et al. (2011), who compare random auditing to two endogenous audit mechanisms similar to the rank-order tournament studied here. They focus on the reporting decision of firms in both their theoretical exposition and in their experimental design, and find that reporting, and hence compliance, is higher with the endogenous audit mechanisms. A key difference is that our study adds an output choice to the reporting decision, enabling us to study the overall efficiency of the audit schemes because output affects net social returns. Similar to how investments in corporate social responsibility are considered as contributions to public goods, lower output choices lead to greater social surplus.

Several experimental papers have tested different regulatory compliance theories, some with dynamic audits based on the compliance history of firms (Alm et al., 1993; Clark et al., 2004; Cason and Gangadharan, 2006). Gilpatric et al. (2012) study a dynamic tournament-based auditing mechanism. Others have used static models, where policy changes such as different penalty rates and tax rates, tax amnesties or changes in audit probabilities are introduced to determine the impact on compliance behavior (for surveys of tax compliance experiments see Alm and McKee (1998) and Torgler (2002)). On the theoretical front, Oestreich (2013) endogenizes output (emissions in his setup) as well as reporting, and shows that optimal output levels are non-monotonic with respect to competition in reporting. Some competition in

reporting leads to fewer emissions compared to the random auditing schemes, however a high level of competition in reporting leads to higher emissions than the random mechanism. In his model competition occurs via a Tullock contest rather than a rank-order tournament. Bayer and Cowell (2009) examine a relative audit theory in the context of tax compliance, where the probability of being audited depends on the firm's observable actions relative to others in the same market. Their theoretical findings suggest that using a relative audit rule generates two dividends--less tax evasion and an efficiency improvement--by making use of the strategic interdependence between firms.

Our paper is also related to the recent but growing research on the behavioral impact of choice observability. Many of these studies document a positive empirical relationship between observability and pro-social choices. Coricelli et al. (2010) explore the impact of public display of identity in tax experiments, where the audit rule is endogenous, with the probability of being audited depending on the median report in the group. Using a within-subject design, they find that displaying the photo of a cheater raises more emotions (as measured by skin conductance responses) and deters cheating. In contrast, Dufwenberg and Muren (2006) find that Dictator game giving declines when identity of donors and recipients is made public.

Making contributions and identity visible has been studied more extensively in public good games, which feature externalities similar to those implemented in the present study. Andreoni and Petrie (2004) show that making contributions of participants visible in conjunction with their identity (using photos) leads to a substantial increase in contributions. In isolation they are not effective, but combining the two features in the information and identification treatment leads to

a 59% increase in contributions to the public good.¹ By contrast, Noussair and Tucker (2007) find that revealing all contribution decisions reduces contributions as compared to revealing none in a repeated game. Dickinson et al. (2015) examine if public punishment, i.e., observing punishment of someone else in a public good setting, leads to an increase in own contributions. While they find that this effect exists, it dissipates over time. Khadjavi et al. (2014) conduct a public good experiment with asymmetric giving and taking action sets and find that making individual contributions public (transparency) helps improve cooperation only when a peer punishment mechanism was introduced. They also find that more transparency surprisingly backfires and leads to a decrease in contributions. In many of the repeated play experiments discussed here, contributions fall over time even with social information. This body of work suggests that while observability certainly has some impact on social behavior in economically-motivated contexts, its influence needs to be investigated more fully in different games and settings.

Our study builds on the results in this literature and makes both identity and actions visible, which is observed to have the most influence on behavior. We explore the impact of visibility in an environment where one of the decisions (the output decision) has a direct negative impact on other group members while the other (the reporting decision) has an indirect but mixed impact. On the one hand, higher reporting indicates greater honesty which can lead to less social stigma.

¹ Rege and Telle (2004) similarly show that making identity and choices public (subjects write down their contributions on a board in full view of the group) leads to higher contributions in a one shot public good game. Samek and Sheremeta (2014) also find that contributions are higher when all contributors are recognized (when photos and names are displayed), and that contributions are affected more by the recognition of the lowest contributor (inducing shame) than by recognition of the highest contributor (inducing prestige). Lopez et al. (2012) find that revealing identity and contribution of one, randomly picked, member out of a five member group significantly increases contributions in a framed public good field experiment in Colombia. Spraggon et al. (2014) investigate the effects of different levels of random revelation and report that while public revelation of individual choices increases contributions in a public good game, in general the difference between revealing three as compared to five subjects' contributions is not significant. This suggests that contributions do not monotonically increase with more revelation.

On the other hand, greater reporting harms other group members in the Tournament mechanism by increasing the likelihood that they are audited. This latter effect could counteract the stigma of being dishonest. Furthermore, in contrast to most other audit experiments, in our research both output and reporting are decision variables, and this helps us isolate the impact of competition on both these aspects of the regulatory environment. Tournaments can improve regulatory compliance by reducing misreporting and by reducing output. Our study makes a novel contribution by identifying output as an additional channel via which regulatory performance should be measured. This is important because the output choices directly determine the net social returns and efficiency for the social dilemma.

Our paper proceeds as follows. In the next section, we develop a theoretical model that provides the framework for our experiments. The experimental design is described in Section 3, followed by the results in Section 4. Section 5 summarizes our findings and discusses some policy implications.

2. Theoretical Model

We present the model in terms of environmental regulation, although as noted in the introduction the results for the alternative enforcement schemes apply to many other applications with voluntary reporting - ranging from income tax reporting to health and safety regulation. Firms must decide on both their level of emissions (e_i) and how many of these emissions to report (r_i) to the regulatory agency. Firms receive a private benefit from their emissions, $B(e_i)$, reflecting, for example, the avoidance of abatement expenditures, where $B'(e_i) > 0$ and $B''(e_i) \leq 0$. Emissions however impose a cost (damages) on all members of society of $C(E)$ where $E = e_1 +$

$e_2 + \dots + e_n$, n is the number of firms, $C'(E) > 0$ and $C''(E) > 0$.² Therefore, by limiting emissions firms contribute to an impure public good. This creates a social dilemma that provides a potential role for social spillovers to motivate decisions.

After choosing e_i the firm must report its emissions to the regulator, however each unit reported incurs a reporting cost (tax) of t . The firm faces a probability of inspection given by $P(e_i, r_i, e_{-i}, r_{-i})$ where e_{-i} and r_{-i} are the emissions and reports, respectively, of other firms. This probability is constant in the case of random auditing, while under the audit tournament for now we simply assume that the probability of inspection is increasing in the degree of under-reporting; i.e. $\frac{\partial P}{\partial e_i} = -\frac{\partial P}{\partial r_i} \geq 0$.³ Firms that are found to have under-reported incur a fine given by $F(e_i - r_i)$ where $F'(e_i - r_i) > 0$ and $F''(e_i - r_i) > 0$.

Thus, the firm's problem is to choose e_i and r_i to maximize expected profit which is given by:

$$\Pi = B(e_i) - C(E) - tr_i - P(e_i, r_i, e_{-i}, r_{-i})F(e_i - r_i)(1)$$

Differentiating (1) with respect to e_i and r_i , respectively, yields the following first order conditions⁴:

$$\frac{\partial \Pi}{\partial e_i} = B'(e_i) - C'(E) - \frac{\partial P(e_i, r_i, e_{-i}, r_{-i})}{\partial e_i} F(e_i - r_i) - P(e_i, r_i, e_{-i}, r_{-i})F'(e_i - r_i) = 0 \quad (2)$$

$$\frac{\partial \Pi}{\partial r_i} = -t - \frac{\partial P(e_i, r_i, e_{-i}, r_{-i})}{\partial r_i} F(e_i - r_i) + P(e_i, r_i, e_{-i}, r_{-i})F'(e_i - r_i) = 0 \quad (3)$$

² This externality $C(E)$ is not needed for any of the results and hypotheses presented here, and it is more relevant for environmental regulation than other tax reporting applications. In particular, $C(E)$ is not in the key comparison between (5) and (6) below. A greater choice for emissions e_i raises firm i 's earnings but imposes a social cost on the other firms through $C(E)$. Other kinds of externalities could be modelled differently.

³ The second partial derivatives of $P(e_i, r_i, e_{-i}, r_{-i})$ are assumed to be zero. We describe later in this section our specific implementation of the tournament.

⁴ The conditions stated above are sufficient to ensure the second order condition holds.

Substituting from (3), using $\frac{\partial P}{\partial e_i} = -\frac{\partial P}{\partial r_i}$, and assuming a symmetric equilibrium, the optimal choice of emissions (e^*) is defined by:

$$B'(e^*) - C'(ne^*) \equiv t \quad (4)$$

As shown in (4), the optimal level of emissions is independent of the enforcement parameters, but is decreasing in the reporting cost (t) and the number of firms (n). This result leads to our first proposition.

Proposition 1: the optimal emissions level is independent of the auditing scheme adopted by the regulator.

Given the optimal choice of emissions (e^*), (3) determines the optimal reporting decision (r^*). If inspections are random, then $\frac{\partial P}{\partial r_i} = 0$ yielding:

$$P(e^*, r_r^*)F'(e^* - r_r^*) \equiv t \quad (5)$$

Intuitively the firm is comparing the reporting cost with the marginal expected fine.

If instead an audit tournament is adopted such that the probability of inspection is increasing in the degree of underreporting then we get:

$$-\frac{\partial P(e^*, r_t^*)}{\partial r_i} F(e^* - r_t^*) + P(e^*, r_t^*)F'(e^* - r_t^*) \equiv t \quad (6)$$

The additional (first) term reflects that the reporting decision now affects the expected fine via a change in the probability of inspection as well as a change in the fine level itself.

Given equivalent overall inspection budgets and a symmetric equilibrium, the inspection probability is identical in both schemes, i.e. $P(e^*, r_r^*) = P(e^*, r_t^*)$. Using this assumption and given that F is convex, a comparison of (5) and (6) reveals that $r_r^* < r_t^*$. That is, reporting is higher under the tournament scheme than with random auditing, given identical emission choices. This leads to our second main hypothesis.

Proposition 2: reporting of emissions is greater under the auditing tournament compared with random auditing.

Let us now turn to the specific operationalization of the two inspection regimes in our setting. If inspections are random, m of the firms are chosen randomly for inspection from the n potential firms, and the regulator then accurately determines their actual emissions e_i . Therefore the probability of inspection is $P(e_i, r_i, e_{-i}, r_{-i}) = m/n$.

For the non-random inspections we model the regulator as conducting a rank order tournament to select the m firms for inspection. This makes an individual firm's probability of inspection dependent on their own and all other firms' emissions and reporting choices. Following the standard approach to such tournaments pioneered by Lazear and Rosen (1981), the regulator observes the firms' performance (emissions) with additive noise, so the estimated emissions for firm i are $\hat{e}_i = e_i + \varepsilon_i$.

In the experiment the distribution of this noise term must be understood by subjects, and we follow the standard practice in the literature to draw the ε_i from a uniform distribution because this is easily explained in experiment instructions. In particular, we employ a uniform distribution for ε_i with support $[-q, +q]$. This has the added benefit of simplifying the marginal impact on the probability of inspection at the symmetric equilibrium where $e_i = e_j = e^*$. As noted by Orrison et al. (2004), this implies that $\frac{\partial P(e_i, r_i, e_{-i}, r_{-i})}{\partial e_i} = -\frac{\partial P(e_i, r_i, e_{-i}, r_{-i})}{\partial r_i} = \frac{1}{2q}$. This is because the uniform distribution with support $2q$ implies that every increase in the amount of underreporting, assuming that others do not change their underreporting, increases the probability of inspection by $1/(2q)$. This can be used in (6), along with the assumption of an equivalent number of inspections as the random inspections case $P(e^*, r_i^*) - m/n$ to determine the equilibrium amount of underreporting.

3. Experimental Design

Decision makers in our setting are firms or individuals who are required to report or disclose an activity. The activity could be of different kinds, such as the level of pollution emitted or compliance towards health and safety standards. The main objective of our experiment is to examine the impact of different enforcement mechanisms on the activity level and on disclosure/reporting rates.

3.1 Decision Making

Subjects participate in one of the four treatments. They interact in groups of five and make decisions in 25 periods. The group members remain in the same groups all through the experiment. In all treatments and all periods, subjects make two decisions. The first is an output decision. This output generates a private benefit, $B(e_i)$, to the subjects, but imposes a cost on everyone in the group, $C(E)$. Table 1 presents the functional forms and the parameters chosen for each of the variables in the experiment. The payoff function for each subject (Private Benefit – Group Cost) reflects the negative externality that their output choice imposes on the group. The group cost depends on the aggregate level of output chosen by everyone and is the same for everyone in the group. The second decision is to choose what level of output to report. For each unit of output they report, subjects pay a reporting cost of 0.70 experimental dollars. After subjects submit their report, the computer (regulator) chooses which of the five members to inspect. Inspection leads to full revelation of actual output.

In the Random treatment only two out of the five members, randomly selected, are inspected every period. If they are inspected and their actual output exceeds their reported output, they are required to pay a fine. The subjects were given a table summarising the fine for each unit of

difference between the actual and reported output. The fine consists of two components: the penalty for under-reporting and the reporting cost the firm avoided by under-reporting.

In the Tournament treatment, once the subjects have submitted their reports, the computer makes an estimate of their output. As the regulator cannot fully observe output before the inspection stage, the estimate includes some randomness. Subjects are informed that the random amount has an equal chance of being any integer number between and including, -17 and 17, with a mean of zero implying that on average estimated output is equal to their actual output. The computer ranks the five members based on their estimated output and their reported output. The two subjects who have the biggest gap between their estimated and their reported output are chosen to be inspected. Table 2 illustrates the selection procedure and fines paid for one specific period in the experiment. As in the Random treatment, at inspection if subjects' actual output is found to exceed their reported output, they have to pay a fine. This penalty function is the same across all treatments.

The Random and the Tournament treatments were conducted with two levels of social information (High and Low) to examine the impact of social observability on decision making.⁵ In all the treatments subjects were given feedback at the end of the period regarding whether they were inspected, their earnings, and the reported output of all other people in their group and whether the group members were inspected. In the High Information treatments, subjects were also given feedback about the actual output and the penalty imposed on the group members who were inspected that period. Importantly they also saw the photographs of all their group members in the High information condition, which were displayed along with their decisions on the

⁵ We chose to examine a Low information treatment instead of a No Information treatment as this represents a more realistic scenario in the regulatory settings we are exploring. Firms are likely to observe at least some information about enforcement concerning their peers, even though this information could be minimal in some cases.

computer screen.⁶ Photographs of group members were not shown in the Low Information treatments. The feedback screen for the High and Low Information treatments is presented in the instructions in the Appendix

All sessions were conducted at the Vernon Smith Experimental Economics Laboratory at Purdue University, using z-Tree (Fischbacher, 2007). All 120 participating subjects were undergraduate students, broadly recruited across different disciplines at the University by email using ORSEE (Greiner, 2004). Although some had participated in other economics experiments, all were inexperienced in the sense that they had never participated in a similar experiment featuring tournaments and social information. The 2x2 experimental design was balanced, with 6 groups of 5 subjects (30 subjects total) assigned to each of the four treatments.

Subjects were randomly assigned to different groups upon arrival. At the beginning of each experimental session an experimenter read the instructions aloud while subjects followed along on their own copy. They participated in 25 periods and this number of periods was common knowledge and announced in the instructions. The framing used in the instructions was context specific. We used the terminology of reporting, inspections, penalty and group costs, and this was maintained exactly consistent across all treatments. While subjects interacted anonymously in 5-person fixed groups, three groups under the same treatment conditions were conducted simultaneously in the laboratory, employing 15 subjects in each session. At the end of the instructions, subjects took a computerised quiz to examine and reinforce their comprehension and understanding of the instructions. The quiz was incentivised and subjects earned \$0.50 for each correct answer. If a subject answered a question incorrectly, her computer presented the correct answer on-screen, with reference to the relevant text in the instructions for explanation.

⁶ Among other things, displaying the photographs of participants captures the reputation effects that managers incur when their firm's pro-social or anti-social actions are revealed publicly. These actions could lead to an increase in prestige or stigma for the firms.

Subjects in the High information treatments were informed that a photograph of subjects' faces would be taken and displayed next to their decisions. The photographs were erased after the session and the subjects were informed about this procedure. This was also explicitly stated in the consent form. None of the subjects raised any objections to their photograph being taken and used in the experiment.

At the end of the experiment subjects participated in a risk task. They received a \$5 endowment and had the option to invest as much as they wanted in \$0.50 increments. The investment returned either zero or three times the amount invested, with equal probabilities (Gneezy and Potters, 1997). This task allows us to elicit subjects' risk preferences and examine if these preferences predict reporting behaviour. Subjects also answered questions relating to demographics such as gender, age and field of study and questions on legal and social norms and motivations for misreporting.

Subject payments comprised of three parts. Firstly, they earned money for each correct answer in the quiz. Secondly, they were paid based on their decisions in five randomly chosen periods in the enforcement task. The experimental dollars from this task were converted to U.S. dollars at a pre-announced 4-to-1 conversion rate. Thirdly, they earned money for their investment decision in the risk task. Subjects' total earnings averaged US\$20.46 each, with an interquartile range of \$13.09 to \$26.59. Sessions usually lasted about 90 minutes on average, including the time taken for instructions and payment distribution.

3.2. Testable Hypotheses

The theoretical framework and the experimental design described above allow us to summarise the following testable hypotheses relating to output and reporting in the four

treatments. First, Proposition 1 and 2 from our theoretical model (Section 2) lead to the following hypotheses regarding the treatment effect of the audit mechanism.

Hypothesis 1a: The output level chosen is the same in the Random and Tournament treatments.

Hypothesis 2a: Misreporting is greater in the Random enforcement treatment as compared to the Tournament treatment.

Second, our parameter values and functional forms (Table 1) generate the following specific predictions regarding the levels of output and misreporting in the two audit treatments.

Hypothesis 1b: The level of output is equal to the theoretical prediction (30) in all treatments.

Hypothesis 2b: Misreporting is equal to the theoretical levels predicted in the Random (21) and the Tournament (8.9) enforcement mechanism.

Finally, while theory suggests that social observability should not affect output and reporting decisions, as discussed in the introduction recent experimental research (such as Coricelli et al, 2010 and Andreoni and Petrie, 2004) suggests that individuals may change their behavior when they are being observed. Hence we expect to find that in the High information treatment, participants will have lower output due to the social cost incurred by higher output and they are also likely to misreport less. This leads to the following two conjectures.

Conjecture 1c: The output level chosen will be lower in the High information treatment than the Low information treatment.

Conjecture 2c: Misreporting will be lower in the High information treatment than the Low information treatment.

The effect of social observability could vary by audit treatments. Higher output choices impose negative externalities on others in the group, so making output observable could lead to participants choosing lower levels of output in both treatments. Making reporting choices visible has no effect on others in the Random treatment, however in the Tournament treatment the effect can be potentially surprising and opposite from output choices. Greater reporting in the Tournament treatment leads to positive externalities for other group members as it reduces their probability of being audited. This could counter the social stigma from creating a negative externality through greater output choices. In the Random treatment, therefore, we expect social observability to improve net social welfare and efficiency, while in the Tournament the effect is ambiguous.

4. Results

In each period subjects choose both output and the amount of output to report. To summarize reporting behavior we construct the following measure: $misreport = output - report$. A positive value indicates underreporting, a negative value over-reporting, and zero indicates accurate reporting.

4.1 Tests of Main Hypotheses

Output

In the experiment, subjects choose a level of output between 0 and 40. Figure 1 plots average output in each treatment over the 25 periods of the experiment. The dashed line shows the theoretical prediction of 30 in all treatments. The figure shows that by the later rounds a clear difference emerges between the two audit treatments, with average output higher when using

random audits compared to the tournament scheme. On the other hand, the information treatment appears to have little impact on output choices.

Table 3 reports average output and misreports in the four treatments. Unless specified otherwise, the unit of observation for all tests is the average at the statistically independent group level, and p-values are from Mann-Whitney rank sum tests. Mann-Whitney tests fail to detect any significant differences in output choices between the two information treatments when averaging across all 25 periods (p-value = 0.75 in Random, p-value = 0.63 in Tournament). Comparing the Random and Tournament treatments, a significant output difference is found with Low Information (p=0.05) but not in the High Information treatments (p=0.15). Pooling over information treatments reveals a statistically significant difference (p=0.02). Considering only output averaged across the last five periods yields similar conclusions.⁷ Hypothesis 1a is therefore partially supported. There is no evidence to support Conjecture 1c.

The efficiency differences across treatments closely mirror the output results. Efficiency in our setting is defined as the total social net benefit generated by the group as a fraction of the maximum social benefit.⁸ Using this measure we find that efficiency in the Random treatment is significantly lower than the efficiency in the Tournament treatment (61% versus 77%; p-value = 0.01). No significant differences were detected between the information treatments. Note that average efficiency in the Tournament treatment is very close to the predicted level of 75%.

The theoretical prediction is that output should equal 30 regardless of the treatment. To evaluate this point prediction we test whether average output in each treatment is different from

⁷ Pooling the information treatments, the difference in output between Random and Tournament is significant (p=0.01) but becomes marginally insignificant when separated by information treatment (p=0.11 in both the Low and High Information treatments). Differences between the two information treatments remain insignificant in the last five periods (p=0.26 in Random; p=0.81 in Tournament).

⁸ The maximum social benefit is 50 when the group total output is 100. At the individually-optimal output choice of 30 (group total output $5 \times 30 = 150$), the net social benefit is 37.5 so predicted efficiency is $37.5/50 = 75\%$.

30 using the Wilcoxon sign rank test. Because no differences were detected between the information treatments we pool the data within the audit treatments. Average output across all periods is marginally significantly different than predicted in both the Random ($p=0.10$) and Tournament ($p=0.08$) treatments. Looking only at the last five periods, the difference is only significant for the random auditing treatment ($p=0.01$) but no longer in the tournament treatment ($p=0.31$).

Thus, output tends to be modestly higher when using Random audits compared to the Tournament audit mechanism and with Random audits is above the equilibrium prediction. Increased experience and learning leads output in the Tournament to tend toward the predicted value. Support for Hypothesis 1b is therefore mixed.

Misreporting

Figure 2 shows the average misreport in each treatment for every period, with the two dashed lines showing the theoretical prediction in the Random (21) and Tournament (8.9) treatments, respectively. The audit Tournament leads to a substantially lower level of misreporting than observed in the Random audit treatment, only about one-third of the amount. Similar to output, no difference between the information treatments is apparent. The figure also shows that misreporting is relatively stable over time in the Tournament audit treatment, but increases considerably with experience with Random auditing.

No significant differences are found across the information treatments in either the Random case or the Tournament when looking across all 25 periods or just the last five periods. Misreporting is significantly different between the Random and Tournament treatments whether disaggregated by information treatment ($p=0.01$ for Low; $p=0.004$ for High) or pooled

across information treatments ($p < 0.0001$). These differences persist in the last five periods. This evidence supports Hypothesis 2a but not Conjecture 2c.

With Random auditing, the predicted amount of misreporting is 21 while it is only 9 in the Tournament case. Pooling the information treatments, average misreporting over all periods is significantly different than these predictions in both the Random ($p = 0.02$) and Tournament treatments ($p = 0.002$). Interestingly, in the Random audit case this difference disappears by the last five rounds, with average misreporting not significantly different from the predicted amount of 21 ($p = 0.24$). The difference however remains in the Tournament treatment ($p = 0.002$). On average, misreporting is only about half of the equilibrium level. This low misreporting is analogous to “overbidding” or over-exertion of effort to “win” the tournament (here, winning means avoiding an audit). Overbidding is observed often in Tullock lottery contests and all-pay auctions, however efforts are usually near equilibrium predictions in rank-order tournaments (Sheremeta, 2013). The observed tournament over-reporting seen here is therefore unusual in the empirical literature.. Evidence for Hypothesis 2b is therefore mixed.

Thus, audit tournaments lead to significantly lower levels of misreporting than random audits. The magnitude of the difference is also meaningful with misreporting about three times greater when using random audits. Finally, experience does not erode the benefit of audit tournaments; on the contrary, experience with the random audit scheme leads to greater levels of misreporting, an effect not observed with the tournament.

4.2 Analysis of Misreporting Behavior

In the previous section we constructed a measure of the average amount of misreporting in the different treatments. In this section we take a more detailed look at misreporting behavior by first investigating if the differences are due to changes in the amount of misreporting or changes

in the proportion of subjects misreporting or both. We then investigate individual patterns in misreporting behavior.

Because the magnitude of under-reporting is censored by a subject's output choice, we construct an alternative measure, *%underreport*, which equals the amount of underreporting as a percentage of the output choice. Consider two subjects who both underreport by 5, while one subject chooses output of 20 and the other 10. The first subject has only underreported by 25% of the maximum possible, while the second subject has underreported by 50%. We also construct an indicator measure of whether a subject misreports at a positive level in a particular period or reports accurately.⁹

Table 4 summarizes these measures and Figure 3 shows the averages both over time and by treatment. The pattern in the top figure, which shows the average underreport as a percentage of the maximum possible, is very similar to the earlier figure on misreporting. Significant differences are found between the *%underreport*, averaged across all periods, comparing Random and Tournament auditing whether disaggregating by information treatment or not ($p < 0.01$ in all cases). These differences also persist into the last five periods. No significant differences are detected across the information treatments or in the proportion of underreporters either across all periods or just the last five periods.

These results indicate that the enforcement mechanism affects the intensive and not the extensive margin. The Tournament auditing mechanism induces subjects to underreport by a smaller amount, rather than reducing the proportion of underreporters. The proportion of underreporters remains roughly constant at a fairly high level in all treatments, at around 70-90% of subjects.

⁹ We exclude over-reports from both of these measures, thus they are measures of *under-reporting*, rather than misreporting more generally. There were 63 instances of over-reporting (i.e. negative values for misreport) or 2% of reports. All but five of these instances were over-reports of 10 or less.

Figure 4 shows the distribution of underreporting across all observations, separately for the two audit treatments but pooling over the (insignificantly different) information conditions. In 40% of observations for the Tournament treatment, underreporting was less than 10% of the maximum possible, which is nearly double the rate for the Random treatment. In the Random audit treatment a similar fraction (just above 20%) underreport by 90% or more of the maximum possible, while this severe underreporting is very rare in the Tournament treatment.

Closer examination of the behavior of those who underreport the maximum possible in the Random treatment reveals that this maximal underreporting increases over time—from less than 9% in the first five periods to over 35% in the last five periods. The participants in this category also chose the highest output level of 40 units most often (frequency of 243 out of 311 cases). This strategy leads to expected profits of approximately 7.20 experimental currency units, which is less than one-third of the expected profits earned from optimal misreporting (about 24). Most subjects (39 out of 60 in the Random treatment) maximally misreport in at least one period. Subjects were also observed to more frequently misreport maximally if they were inspected in the previous period, consistent with the gambler's fallacy of misunderstanding serially uncorrelated random events.¹⁰

While the proportion of subjects misreporting in a given period is high (70% or more), a natural question to ask is whether most subjects usually misreport. That is, are some individuals (nearly) always honest or dishonest? Figure 5 shows the distribution of the number of periods (out of 25) of honest reporting (i.e. report=output) for each of the 120 subjects. The figure shows that 17% of subjects misreported in every period (i.e. number of honest periods=0) and that the majority of subjects misreported in most periods (i.e. 58% of subjects were honest in only 3 or

¹⁰ This result comes from a random effects probit regression explaining the likelihood of maximal misreporting. Errors were clustered by group. The coefficient on lagged inspections is significantly positive ($p < 0.001$), while the time trend is also significantly positive ($p < 0.001$).

fewer periods). No one reported honestly in every single period; two people were honest in 24 periods and one in 21 periods.

Finally, consider the amount of unreported output discovered by the two inspecting mechanisms. In the Random treatment, the average amount unreported by the inspected subjects (15.0) is, by design, not significantly different from the average amount unreported by the uninspected (14.9). As the tournament mechanism targets those more likely to be noncompliant, in this treatment the average amount unreported by the inspected subjects (5.7) is much greater than the average amount unreported by the uninspected (3.3). Note that the lower level of detected misreporting in the Tournament mechanism (5.7 per inspected firm) compared to the random mechanism (15.0) should not be interpreted as poorer audit performance. The lower detected misreporting follows directly from this mechanism's superior ability to motivate greater reporting.

4.3 Individual Motivations

Subjects completed a post-experiment questionnaire intended to reveal information about their motivations for whether or not to obey a law in general, as well as for the choices made during the experiment. Table 5 shows the responses in each category for two key types of motivations: the role of social stigma, and the role of detection/penalties. Whether considering general law obedience or experimental choices, the fear of penalties is rated considerably more important than social pressure factors.

For example, 19% of subjects rated fear of social stigma as an extremely important motivation for obeying laws, compared to 55% rating fear of financial penalties as extremely important. A Wilcoxon signed-rank test of differences between these responses is highly significant ($p < 0.001$; with 54% of subjects reporting a higher importance for financial penalties

and only 3% the other direction). A similar difference is observed in motivations for reporting in the experiment, with 30% rating financial penalties as extremely important compared to only 7% highly motivated by social concern. Again, this difference in responses is highly significant (signed-rank test $p < 0.001$, with 68% of subjects rating penalty as more important than social stigma and only 8% the opposite direction).

Finally, as expected social concerns were significantly less important in the experimental setting than for real world decisions. However nearly half of the subjects were barely concerned with how their choices appeared to others in their group, rating social appearance as “not at all” or “slightly” important. This is consistent with our findings regarding the lack of a behavioral influence for the two information treatments. No differences in individual motivations were detected between treatments.

4.4 Subject Characteristics, Output and Reporting Choices

Table 6 presents regression results using panel regressions that employ random effects at the individual level and clustering at the group level. We estimate three different equations using output, misreport, or %underreport as the dependent variable. The first set of regressions only include experimental variables: indicators for the two treatments (High information and Tournament), separately and interacted with each other, a trend variable ($1/\text{period}$), and the interaction between Tournament and trend, as explanatory variables. These models, shown in the odd-numbered columns in Table 6, confirm our earlier findings regarding treatment differences. The indicator for High information is never significant either separately or when interacted with Tournament. On the other hand, Tournament significantly reduces output, the amount misreporting and the degree of underreporting. The effect of Tournament is not only statistically significant but large in magnitude. The negative estimated coefficient on the trend

variable (1/Period) indicates that both output and misreporting increase over time. However, as the interaction between Tournament and trend indicates, this trend is driven by behavior in the Random treatments. Indeed, in the Tournament treatments there is no significant trend over time in either measure of misreporting ($p=0.660$ Model 3; $p=0.236$ Model 5) and although output does increase over time ($p=0.042$ Model 1), the increase is only around half of that in the Random treatments. These findings are consistent with the results reported earlier.

The second set of regressions contain various demographic control variables such as gender (*male*), place of birth (*US born*) and having a GPA of 3.5 higher (*high GPA*) as well as a measure of risk preference (*Amount Invested* in the risk task), and indicators for religion (*no religion* and *Christian*).¹¹ They also include indicators for whether social stigma and penalties were rated as extremely important motivators for experimental choices and an indicator for subjects who had considerable past experience in experiments.

The first thing to note is that including these control variables does not alter any of our conclusions about treatment effects. Second, the control variables have little explanatory power for output choices, although interestingly subjects who were extremely motivated by social stigma concerns chose a significantly smaller output amount. Third, males, less risk averse, and more experienced subjects misreport more, while high achieving subjects misreport less. These results are consistent whether using misreport or %underreport as the dependent variable. Subjects who indicate they have “no religion” or are “Christian” misreport more than those from

¹¹ Slightly over half of the subjects were male (54%), with 59% born in the US and 34% born in Asia. The most common majors were engineering (27%), science (15%), management or business (11%), and economics (8%). The average amount subjects invested in the risk elicitation task was \$3, with dual modes of \$2.50 and \$5.00; men invest significantly more than women (\$3.45 versus \$2.59, $p=0.001$, Mann-Whitney test). A quarter (26%) of participants state they have no religion, while 44% identify as Christian.

other religions. Motivational factors expressed in the post-experiment questionnaire were insignificant determinants of misreporting.¹²

5. Conclusion

Designing appropriate audit mechanisms that can improve compliance outcomes cost-effectively is critical for improving regulatory efficiency. In this paper we examine two audit mechanisms. While one is based on a simple random audit system and is implemented often in the field, the other has its roots in competition and tournaments. We compare the endogenous tournament mechanism with the random audit system while also varying orthogonally the social observability of actions, and focus on two main performance measures - output and reporting.

Although optimal output levels are theoretically independent of audit schemes, experimentally, output in the tournament treatment is more socially optimal than in the random audit treatment. While these treatment differences are modest, and output is generally close to noncooperative Nash equilibrium levels, efficiency is however substantially higher in the tournament treatment. Thus using competition in enforcement can have benefits that extend beyond simply encouraging more honest reporting. The enforcement tournament appears to generate behavioral spillovers from the reporting stage to the output stage that are not predicted by theory. In terms of reporting, both theoretically and experimentally, the tournament mechanism performs much better than random audits. Tournament competition to avoid being audited reduces misreporting by 60 to 80 percent on average, which is even larger than the reduction theoretically predicted.

¹² We also included additional regressors for the three most common majors and indicators for the importance of religion in daily life (very important and somewhat). As these variables were never significant, either individually or jointly, we do not report them here.

In our framework individuals impose direct negative externalities on others if they choose higher output levels. In spite of this design feature, making actions observable does not have any significant effect on output or reporting in either of the treatments. This is contrary to a recent and rapidly-growing body of literature indicating that observability promotes pro-social behavior. It is possible that when information on reports is available in addition to output, most individuals focus on reports. The reporting decision does not directly affect others in the group in the random scheme, but in the tournament mechanism, greater misreporting actually *benefits* other group members by lowering their likelihood of audit (*ceteris paribus*). This could lead to offsetting incentives, counteracting any social stigma that might exist for being dishonest. Moreover, most individuals misreport at least some of the time in our experiment so it would seem that misreporting is generally acceptable in this context. Thus being exposed as a dishonest reporter does not appear to go against a strong social norm. Individuals may also feel compelled to risk misreporting to try to overcome the high group cost being imposed on them due to the negative externality created by higher output choices. Providing social information about output and reporting choices hence can lead to several subtle effects in the regulatory area. This suggests that social observability, such as through public disclosure of toxic chemical releases (e.g., US EPA, 2014), may have limits as an effective regulatory tool.

In contrast to most of the previous research on auditing which has focused on reporting behavior, our study can provide insights about the impact of different audit mechanisms on both output and reporting. In many contexts, including many important environmental- or health-related applications, output and reporting are important criteria for the evaluation of audit systems. While reporting is perhaps more relevant in the taxation area, complying with output standards is critical to ensure that outcomes more closely adhere to socially optimal standards.

Our findings suggest that tournament audits are more effective than random audits in improving both output and reporting levels, and the differences between them increase substantially over time.

There is some evidence that policymakers are moving towards non-random audit schemes. For example, federal tax authorities in the U.S. are much more likely to audit individuals who have high income (IRS, 2014, Table 9b). There is however little research on whether this is an appropriate target group, and if this is due to a belief that high income individuals are more likely to cheat or because noncompliant high earners allow the authorities to collect more revenue from unpaid taxes and fines.¹³ In other regulatory areas, it may be difficult to determine appropriate targets for regulatory authorities and to evaluate how effective targeting would be. Findings from our research indicate that providing appropriate incentives to comply could be important in such situations. The tournament audits use a competitive and endogenous selection mechanism that relies on relative perceived performance amongst regulated agents. The resulting incentives lead to improved efficiency by both reducing negative-externality generating output and by increasing truthful reporting.

¹³ While our study does not explicitly vary firm size or the marginal benefits incurred by firms, the tournament mechanism can address heterogeneity amongst firms in different ways. For example, regulators could put firms in different groups, based on industry and regions, and apply the audit tournament to smaller groups of relatively homogenous firms.

References

- Andreoni, J., and R. Petrie. 2014. "Public Goods Experiments Without Confidentiality: A Glimpse into Fund-Raising." *Journal of Public Economics*, 88: 1605-1623.
- Alm, J., and M. McKee. 1998. "Extending the Lessons of Laboratory Experiments on Tax Compliance to Managerial and Decision Economics." *Managerial and Decision Economics*, 19, 259-75.
- Alm, J., M. Cronshaw, and M. McKee. 1993. "Tax Compliance with Endogenous Audit Selection Rules." *Kyklos*, 46, 27-45
- Bayer, R., and F. Cowell. 2009. "Tax Compliance and Firms' Strategic Interdependence", *Journal of Public Economics*, 93, 1131-1143.
- Cason, T. N., and L. Gangadharan. 2006. An Experimental Study of Compliance and Leverage in Auditing and Regulatory Enforcement." *Economic Inquiry*, 44, 352-366.
- Clark, J., L. Friesen, and A. Muller. 2004. "The Good, the Bad and the Regulator: an Experimental Test of Two Conditional Audit Schemes." *Economic Inquiry*, 42, 69-87.
- Coricelli, G., M. Joffily., C. Montmarquette and M. C. Villeval. 2010. "Cheating, Emotions, and Rationality: An Experiment on Tax Evasion". *Experimental Economics*, 13, 226-247.
- Dickinson, D., E. G. Dutcher and C. Rodet. 2015. "Observed Punishment Spillover Effects: A Laboratory Investigation of Behavior in a Social Dilemma", *Experimental Economics*, forthcoming.
- Dufwenberg, M., and A. Muren. 2006. "Generosity, Anonymity, Gender," *Journal of Economic Behaviour and Organization*, 61, 42-49.
- Gilpatric, S., C. Vossler and L. Liu. 2012. "Using Competition to Stimulate Regulatory Compliance: a Tournament-Based Dynamic Targeting Mechanism", Working Paper.
- Gilpatric, S., C. Vossler and M. McKee. 2011. "Regulatory Enforcement with Competitive Endogenous Audit Mechanisms.", *RAND Journal of Economics*, 42(2), 292-312.
- Internal Revenue Service. 2014. *Data Book, 2013*. Publication 55B. Washington, DC. Available at <http://www.irs.gov/pub/irs-soi/13databk.pdf> (accessed 19 September 2014).
- Khadjavi, M., A. Lange and A. Nicklisch. 2014. "The Social Value of Transparency and Accountability: Experimental Evidence from Asymmetric Public Goods Games" WiSo-HH Working Paper No. 12, University of Hamburg.
- Kitzmueller, M. and J. Shimshack. 2012. "Economic Perspectives on Corporate Social Responsibility." *Journal of Economic Literature*, 50, 51-84.

- Konar, S., and M. Cohen. 2001. "Does the Market Value Environmental Performance?" *Review of Economics and Statistics*, 83, 281-289.
- Lazear, E. and S. Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy*, 89(5), 841-864.
- Lopez, M., J. Murphy., J. Spraggon., and J. Stranlund. 2012. "Comparing the Effectiveness of Regulation and Pro-social Emotions to Enhance Cooperation: Experimental Evidence from Fishing Communities in Colombia." *Economic Inquiry*, 50, 131-142.
- Noussair, C., and S. Tucker. 2007. "Public Observability of Decisions and Voluntary Contributions in a Multi-Period Context." *Public Finance Review*, 35, 176-198.
- Oestreich, A. M. 2013. "Firms' Emissions and Self-Reporting under Competitive Audit Mechanisms" Available at SSRN: <http://ssrn.com/abstract=1985757> or <http://dx.doi.org/10.2139/ssrn.1985757>.
- Orrison, A., A. Schotter and K. Weigelt. 2004. "Multiperson Tournaments: An Experimental Examination," *Management Science*, 50, 268-279.
- Rege, M., and K. Telle. 2004. "The Impact of Social Approval and Framing on Cooperation in Public Good Situations." *Journal of Public Economics*, 88, 1625-1644.
- Samak, A., and R. Sheremata. 2014. "Recognizing Contributors: An Experiment on Public Goods", *Experimental Economics*, 17, 673-690.
- Sheremeta, R. 2013. "Overbidding and Heterogeneous Behavior in Contest Experiments." *Journal of Economic Surveys*, 27, 491-514.
- Spraggon, J., L. Sobarzo and J. Stranlund. 2014. "A Note on Stochastic Public Revelation and Voluntary Contributions to Public Goods", Working paper.
- Torgler, B. 2002. "Speaking to Theorists and Searching for Facts: Tax Morale and Tax Compliance in Experiments." *Journal of Economic Surveys*, 16 (5), 657-83.
- U.S. Environmental Protection Agency. 2014. *Toxics Release Inventory National Analysis Overview*. Available at http://www2.epa.gov/sites/production/files/2014-01/documents/complete_2012_tri_na_overview_document.pdf (accessed 19 September 2014).

Table 1: Parameters and Equilibrium Predictions

Notation/Functional Form	Definition	Parameters
N	Number of Firms	5
e_i	Output(emissions) of firm i	Output ranges from 0 to 40 units.
$B(e_i) = e_i$	Benefit of firm i from Output(emissions) chosen	Each unit of output generates \$1 for the firm
$C(E) = E^2/1000$	Cost of sum of output chosen by all firms	Depends on output chosen by the 5 firms
t	Per unit reporting cost	0.7
$P(e^*, r_r^*) = P(e^*, r_t^*)$ [-q, +q]	Probability of Inspection For Tournament Audit: The noise added to their actual output.	0.4 [-17, +17]
$F(e_i - r_i) = (e_i - r_i)^2/40 + t(e_i - r_i);$	The total fine for underreporting. Firms pay both a (convex) penalty for under-reporting plus the reporting cost (taxes) they avoided.	
Equilibrium Predictions	Random	Tournament
Optimal choice of output (e^*)	30	30
Optimal choice of reporting (r^*)	9	21.1

Table 2: Example of Tournament Mechanism (Group 1, Period 12, Session 140310_1535)

Subject	Output	Report	Random Draw	Estimated Gap	Inspected?	Fine Paid	Profit
1	35	28	10	17	Yes	6.125	12.375
4	30	25	5	10	No	0	15.6
7	20	15	-8	-3	No	0	12.6
10	25	25	10	10	No	0	10.6
13	20	12	7	15	Yes	7.2	7.5

Table3: Average Output and Misreport by Treatment (Standard Deviation)

Treatment		Average Output		Average Misreport	
Audit	Information	All Periods	Last Five Periods	All Periods	Last Five Periods
<i>Random</i>	<i>Low</i>	31.1 (3.8)	32.5 (3.1)	14.8 (7.1)	17.4 (6.9)
<i>Random</i>	<i>High</i>	31.6 (2.5)	34.8 (3.1)	15.1 (3.0)	20.8 (3.7)
<i>Tournament</i>	<i>Low</i>	28.4 (1.7)	28.7 (3.6)	4.0 (1.9)	3.8 (2.2)
<i>Tournament</i>	<i>High</i>	28.5 (3.8)	28.1 (6.5)	4.5 (2.2)	5.1 (2.4)

Table4: Average Degree of Underreporting and Proportion of Underreporters by Treatment (Standard Deviation)

Treatment		Average % Underreport		Average Proportion Underreporters	
Audit	Information	All Periods	Last Five Periods	All Periods	Last Five Periods
<i>Random</i>	<i>Low</i>	43.2 (17.6)	48.7 (17.8)	0.81 (0.12)	0.83 (0.14)
<i>Random</i>	<i>High</i>	44.1 (6.7)	55.2 (8.8)	0.80 (0.09)	0.87 (0.12)
<i>Tournament</i>	<i>Low</i>	15.5 (5.6)	15.1 (7.1)	0.76 (0.13)	0.84 (0.13)
<i>Tournament</i>	<i>High</i>	17.8 (10.8)	19.9 (12.5)	0.71 (0.22)	0.75 (0.21)

Table5: Motivations for Obeying Laws and Misreporting in the Experiment

<i>% of responses in each category</i>	<i>Not at all important</i>	<i>Slightly important</i>	<i>Somewhat Important</i>	<i>Important</i>	<i>Extremely Important</i>
<i>General Law Obedience</i>					
Fear of social embarrassment / stigma for breaking law	6	12	30	33	19
Fear of financial or other legal penalties for breaking law	2	3	6	35	55
<i>Accurate reporting in experiment</i>					
Concern about how my report and output choice appears to others in my group	30	16	24	23	7
Fear detection/financial penalties	0	5	18	46	30

Notes: *Question 14: Please indicate how you would rate the importance of the following personal motivations for your decision to obey or not obey a law. Question 18: Importance of the following motivations for experimental choices*

Table 6: Panel Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
	output1	output2	misreport1	misreport2	%underreport1	%underreport2
trend (1/Period)	-9.1990*** (1.7436)	-9.0369*** (1.7699)	-12.9536*** (2.3446)	-12.5113*** (2.3810)	-27.1893*** (6.5858)	-26.2914*** (6.6921)
High Information	0.4884 (1.7215)	-0.1010 (1.6644)	0.3824 (2.9437)	0.5520 (2.6611)	0.7097 (7.2071)	2.4733 (6.4425)
Tournament	-3.4923** (1.7771)	-4.0082** (1.9499)	-12.7141*** (3.1062)	-12.2828*** (2.9143)	-32.3833*** (7.7595)	-29.2637*** (7.2776)
High Information * Tournament	-0.4007 (2.3417)	-0.2014 (2.3808)	0.1209 (3.1392)	-0.8778 (3.0469)	1.6769 (8.5755)	-2.9462 (7.9808)
Tournament * trend	5.1929** (2.6370)	5.1867** (2.6431)	12.6275*** (2.4588)	12.6106*** (2.4566)	30.5443*** (7.1686)	30.4912*** (7.1654)
male		0.0166 (0.8322)		1.8006* (0.9674)		6.6474** (2.7910)
USborn		0.3914 (1.3861)		-1.4116 (1.3095)		-5.5719 (3.5770)
high GPA		0.1809 (1.5818)		-3.0052** (1.4399)		-7.3297* (3.8441)
experienced subject		-2.2747 (1.4698)		3.6894** (1.5509)		19.0492*** (4.7931)
noreligion		1.0803 (1.0412)		2.5828*** (0.8442)		7.3705*** (2.8595)

Christian		-0.0344 (1.1580)		2.1166** (0.9007)		6.3337*** (2.3506)
AmountInvested		0.3563 (0.2631)		0.9722*** (0.2388)		1.9164*** (0.5265)
stigma extremely important experimental motivation		-4.7976** (2.2414)		-2.1508 (2.3663)		-4.7390 (6.0274)
penalty extremely important experimental motivation		-1.2109 (1.2899)		-1.1106 (1.2584)		-3.7173 (3.0809)
Constant	32.5121*** (1.5762)	33.6839*** (2.1416)	16.7319*** (3.0174)	16.3753*** (3.1040)	47.3654*** (7.4535)	43.3151*** (8.0144)
Observations	3000	3000	3000	3000	2937	2937
Subjects	120	120	120	120	120	120
Wald test of model (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Standard errors in parentheses based on individual subject random effects and clustering at the group level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: Average Output by Treatment

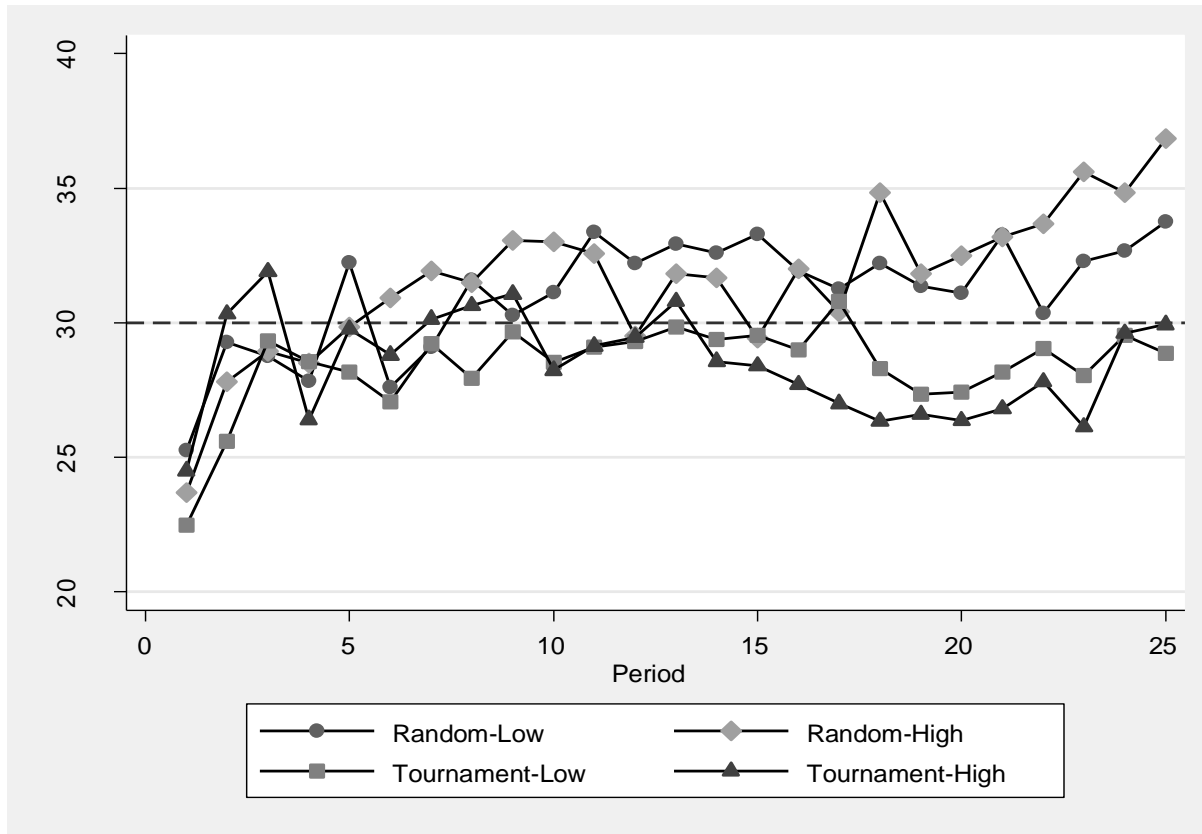


Figure 2: Average Misreport by Treatment

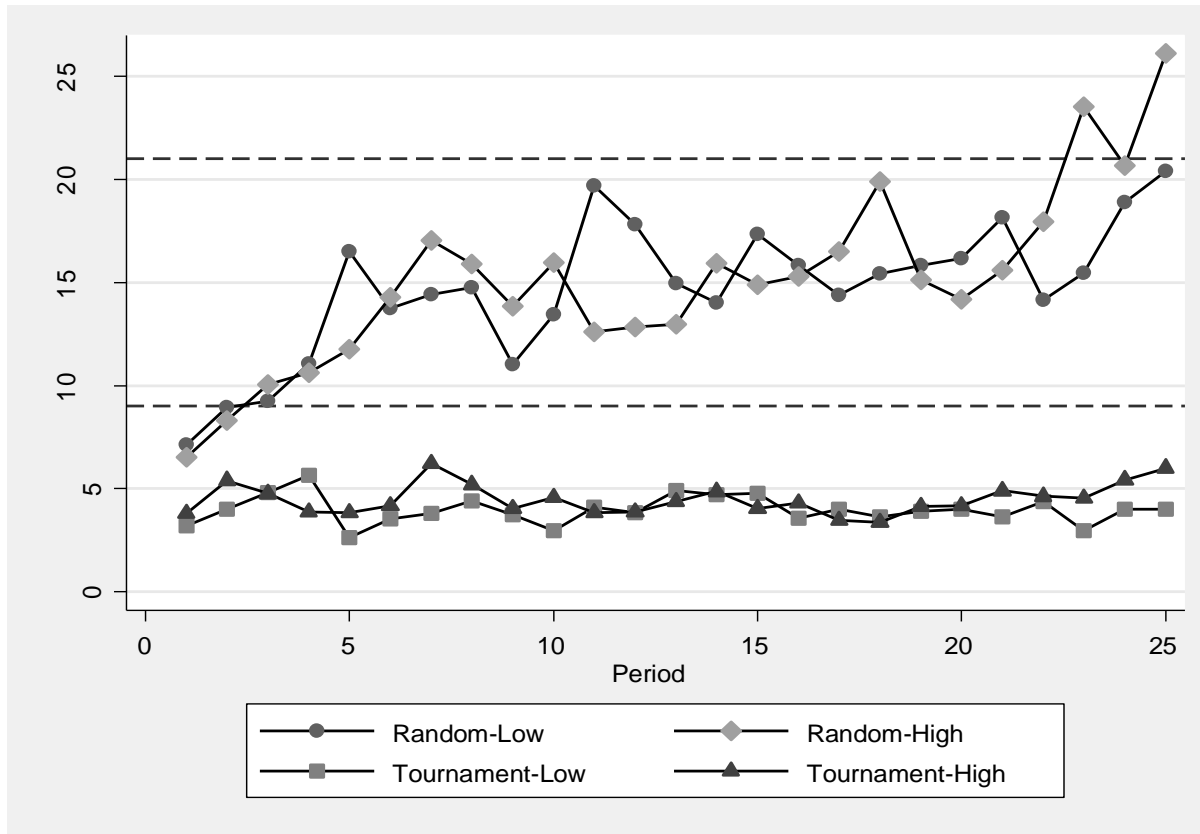
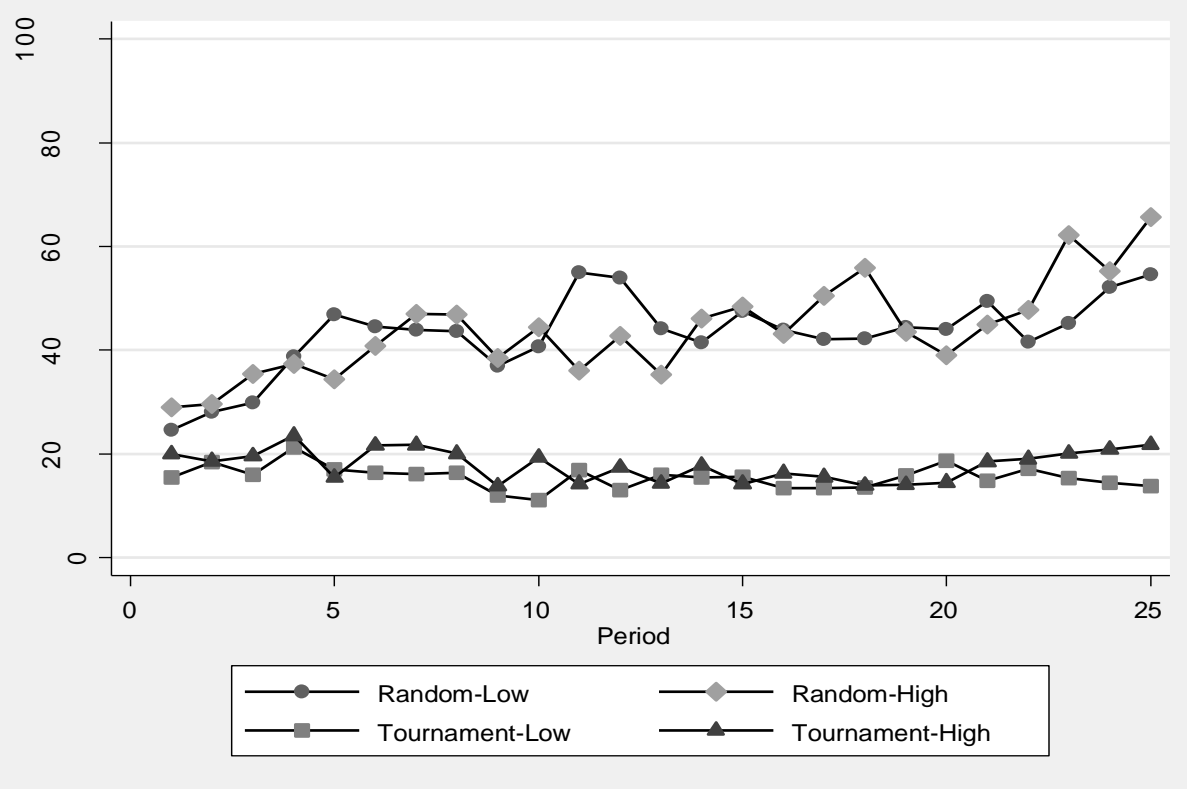


Figure 3: Average Degree of Underreporting and Proportion of Underreporters by Treatment



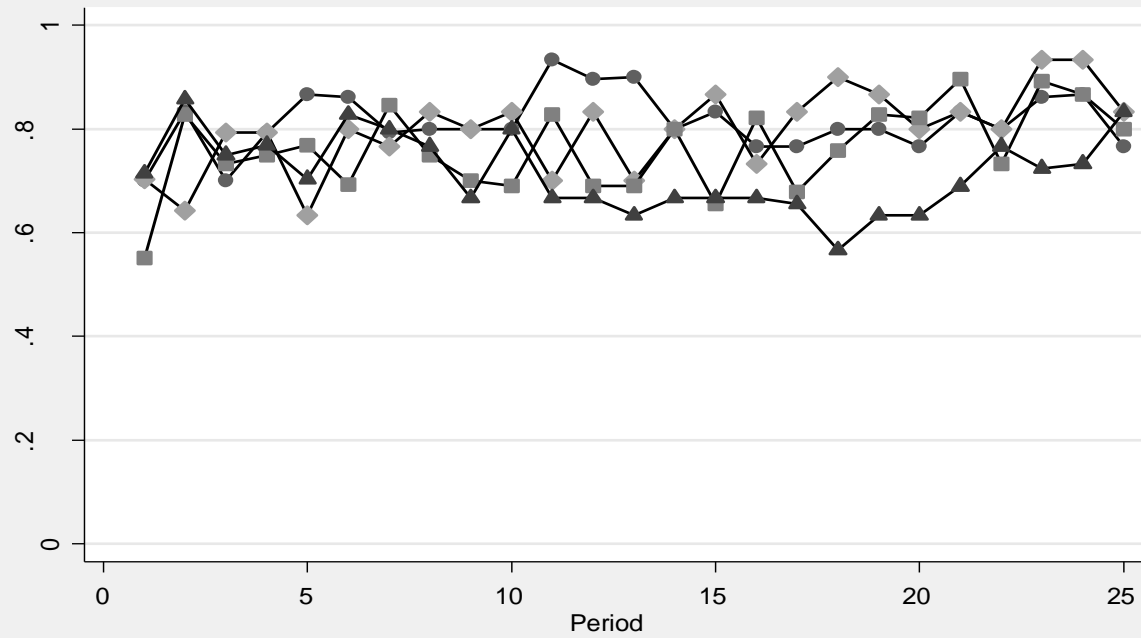


Figure 4: Distribution of Degree of Underreporting by Audit Treatment

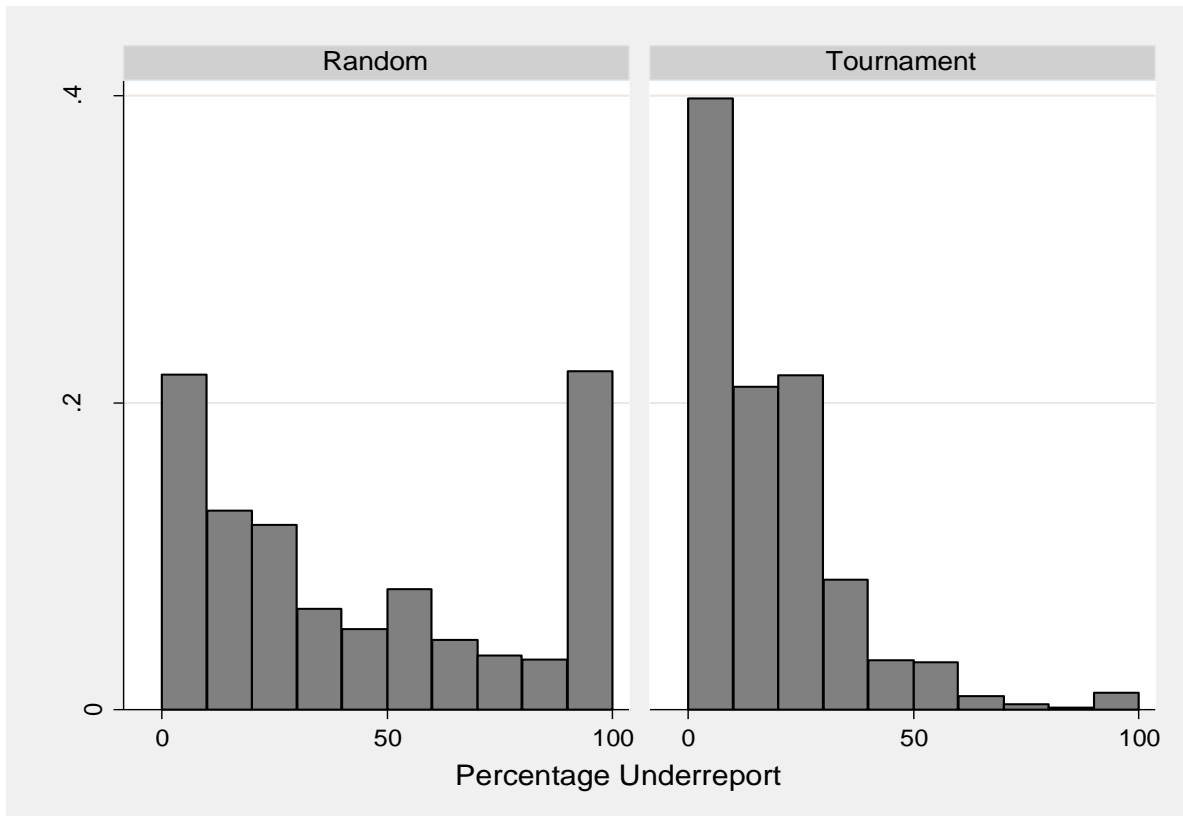


Figure 5: Individual Subject Frequencies of Number of Periods of Honest Reporting

