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

March 2016

Abstract

This paper examines the effectiveness of traditional regulatory schemes and newly emerging social information schemes for achieving compliance. Our experiment focuses on 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 underreporting. To study the effects of social information, the experiment varies the observability of identity, output, and compliance decisions. Optimal output is theoretically independent of the auditing scheme, but equilibrium reporting is higher under the Tournament mechanism than Random auditing. Experimental findings are broadly consistent with the theoretical predictions for reporting, but deviate modestly 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

*Corresponding author. Department of Economics, Purdue University, 100 S. Grant Street, West Lafayette, IN 47907-2076, U.S.A.

1. Introduction

Policy makers employ various approaches to encourage individuals to consider the impact of their actions on others. Some of these approaches are formal and rely on monitoring and auditing, with penalties for observed non-compliance. Others can be more informal and allow the use of information mechanisms to influence behavior. Our goal in this paper is to examine the effectiveness of audit mechanisms and information observability and the interaction between these different approaches. This is important to study because information policies are often implemented in conjunction with traditional enforcement measures, perhaps due to budget constraints faced by enforcement agencies. For example, environmental regulators in the U.S. are committed to using "Next Generation compliance that takes advantage of new information and monitoring technologies," in addition to traditional enforcement programs (US EPA, 2014a, p.39). As budgets available for traditional enforcement tighten while the costs of informal enforcement via information disclosure fall due to the rising prevalence of social media, it is important to know whether and how these different approaches complement each other. To provide insight about the potential value of these new information-based approaches as either beneficial or detrimental for traditional regulations, it is critical to better understand how they interact.

Auditing is a traditional enforcement approach that helps ensure that the regulatory programs instituted by governments are effective and that individuals and organizations are appropriately compliant with regulations. Beyond auditing of financial accounts for tax reporting, this can include inspections for compliance with health, safety and environmental regulations. Auditing however is expensive and consumes scarce regulatory budgets, so substantial benefits can arise from audit mechanisms that improve compliance cost-effectively.

In particular, audit mechanisms that choose whom to audit randomly may spend limited resources auditing mostly compliant agents, so regulators may prefer to introduce targeting schemes that are more likely to apprehend the biggest or most frequent violators. There is some evidence of growing use of such non-random audit schemes (e.g., IRS, 2014). This paper examines one such mechanism 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 interest in a competitive mechanism stems from the notion that tournaments can be efficiency enhancing and audits based on relative performance could help improve social outcomes cost effectively.

The information mechanism that we investigate in this paper focuses on whether increasing the visibility of actions can lead to improved compliance. Firms and individuals are often motivated by their public image. For example, they may suffer disutility when being revealed as violators of social norms (Perez-Truglia and Troiano, 2015; Samak and Sheremeta, 2014; Coricelli et al., 2010; Andreoni and Petrie, 2004) or face adverse market valuations when poor compliance is publicly revealed (Konar and Cohen, 2001). Therefore, making actions and audit outcomes 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 public disclosure are mostly negligible.

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. Especially for innovative mechanisms, that combine different approaches, it is difficult to find appropriate counterfactuals, making empirical evaluation based on field data challenging. 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 others in society. These 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. In the model 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, they 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 improve compliance, 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 and 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 underreporting. Hence firms have an incentive to compete with each other in terms of compliance, since lower underreporting 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 by an equivalent amount 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. The social visibility of actions and identities may influence output and reporting choices by increasing the stigma associated with choices that deviate from the social norm. Further, the impact of visibility could differ across audit treatments. We hypothesize that increased compliance will occur in the audit tournament, and based on previous experimental results summarized in Section 2, we expect that increased information disclosure about reporting and compliance may also increase compliance. The effects of these enforcement and disclosure margins may reinforce each other or perhaps unexpectedly conflict, so we also include an interaction treatment where they are both present.

Our study contributes by investigating how the resources deployed in traditional enforcement could be used more effectively by using competition, as well as studying whether formal regulatory rules can be complemented through informal policies such as the disclosure of past compliance. Furthermore previous studies have considered information disclosure in non-competitive environments so our experiment is novel in considering information in a competitive environment where the effects are unexplored and uncertain. The interaction between these two treatment dimensions can provide significant insights to both researchers and policy makers. Technological advancement has made information policies viable and cheaper while fiscal constraints in many jurisdictions have made traditional auditing more challenging. Examining the effectiveness of these two approaches for promoting compliance is hence very timely.

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 the experiment are broadly consistent with the theoretical predictions for reporting, but deviate modestly 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, contrary to recent findings in the literature discussed in the next section.

2. Related Literature

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 in the present study lead to greater social surplus.

Several experiments have examined different regulatory compliance theories, some of these include dynamic audits based on the compliance history of firms (Alm et al., 1993; Clark et al., 2004; Cason and Gangadharan, 2006). Gilpatric et al. (2015) also study a dynamic model and focus on a 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)). Oestreich (2015) endogenizes output (emissions in his setup) as well as reporting, and shows theoretically that optimal output levels are non-monotonic with respect to competition in reporting. The competitive audit mechanisms he models involve either a Tullock contest or an all-pay auction 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. 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 recent 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 reporting 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.[1] Perez-Truglia and Troiano (2015) provide field evidence that even though making identities of tax delinquents' public is less effective than financial penalties for increasing revenue, public shaming has a significant role to play in increasing the speed of debt repayment. 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 (59%) increase in contributions.[2] 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

[1] Coricelli et al. (2014) find that public display of identity only reduces cheating in situations where individuals can be forgiven and reintegrated in the group. In the absence of this option individuals do not change behavior as they do not experience more shame with each subsequent cheating decision. Once their reputation is lost they have no reason not to be dishonest.

[2] 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. (2015) 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.

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. These mixed results indicate that observability certainly has some impact on social behavior in economically-motivated contexts, but 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 signals greater truthfulness which can lead to less social stigma. 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 revealed as untruthful. Furthermore, in contrast to most other audit experiments, in our research both output and reporting are decision variables, and this helps isolate the impact of competition on both these aspects of the regulatory environment. We find that 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 in the social dilemma.

3. Theoretical Model

To motivate our experimental design, in this section we develop a simple model. The setup of the model is as follows. Firms produce output, and the level of output must be reported to the regulatory agency. However reporting is costly with each unit reported incurring a cost. The regulator receives the report from the firm, as well as a noisy signal regarding the firm's actual output. The regulator cannot directly observe actual output unless an inspection is conducted. Inspections perfectly reveal the firm's actual output and thus the accuracy of their report, but the regulator only has a limited number of inspections available. Depending on the experimental treatment, these limited inspections can be allocated either randomly or by using a rank-order tournament where the firms believed to have the highest degree of underreporting are inspected. Firms that are revealed to have underreported their output face a penalty that increases in their degree of underreporting.

More precisely, firms must decide on both their level of output (y_i) and how much of this output to report (r_i) to the regulatory agency. Firms receive a private benefit from their output $B(y_i)$, where $B'(y_i) > 0$ and $B''(y_i) \leq 0$. Output however imposes a cost on all members of society of $C(Y)$ where $Y = y_1 + y_2 + \cdots + y_n$, n is the number of firms, $C'(Y) > 0$ and $C''(Y) > 0$.[3] In the context of environmental regulations, for example, output reflects emissions, the private benefit is the avoidance of abatement expenditures, and the cost on society is

[3] This externality $C(Y)$ is not needed for any of the results and hypotheses presented here, and it is more relevant for environmental regulation than other applications, such as tax reporting or health and safety regulations. (For example, common pool resources are usually modelled such that actions of one agent create an externality for the others in the group.) In particular, $C(Y)$ is not in the key comparison between (5) and (6) below. A higher choice for output y_i raises firm *i*'s earnings but imposes a social cost on the other firms through $C(Y)$. Other kinds of externalities could be modelled differently.

emission damages. Therefore, by limiting emissions firms contribute to an impure public good. In general, we are interested in any situation in which actions that benefit an individual firm impose negative externalities on others. This creates a social dilemma that requires regulation to improve efficiency, and also provides a potential role for social spillovers to motivate decisions.

After choosing y_i the firm must report its output to the regulator, and each unit reported incurs a reporting cost (tax) of t. Following the standard approach to rank order tournaments pioneered by Lazear and Rosen (1981), the regulator observes the firms' output with additive noise, so the estimated output for firm i is $\hat{y}_i = y_i + \varepsilon_i$ where ε_i is a uniformly distributed random variable with support [-q, +q].

The firm faces a general probability of inspection given by $P(y_i, r_i, \boldsymbol{y_{-i}}, \boldsymbol{r_{-i}})$ where $\boldsymbol{y_{-i}}$ and $\boldsymbol{r_{-i}}$ are *(n-1)* dimensional vectors comprising the emissions and reports, respectively, of the other firms. The regulator is only able to conduct m inspections where $m < n$. We consider two ways of allocating these limited inspections. First, if firms are randomly chosen for audit, then each firm has a probability of inspection $P(y_i, r_i, \boldsymbol{y_{-i}}, \boldsymbol{r_{-i}}) = m/n$. Alternatively, the regulator may use a rank order tournament where the m firms with the highest estimated degree of underreporting, $\hat{y}_i - r_i$, are inspected. In this latter case, each firm's probability of inspection is (weakly) increasing in its degree of underreporting; i.e. $\frac{\partial P(y_i, r_i, \boldsymbol{y_{-i}}, \boldsymbol{r_{-i}})}{\partial y_i} = -\frac{\partial P(y_i, r_i, \boldsymbol{y_{-i}}, \boldsymbol{r_{-i}})}{\partial r_i} \geq 0$.[4]

Firms that are found to have underreported incur a fine given by $F(y_i - r_i)$ where $F'(y_i - r_i) > 0$ and $F''(y_i - r_i) > 0$. Note that unpaid reporting costs are included in our fine function. The convex fine yields an interior solution for reporting when audits are random and perfect.

The firm's problem is to choose y_i and r_i to maximize expected profit which is given by:

[4] Our model is similar to that of Gilpatric et al. (2011) although they do not model the output choice. Other differences are that in their model audits are imperfect and the penalty function is linear. They also consider general functional forms for the errors.

$$\Pi = B(y_i) - C(Y) - tr_i - P(y_i, r_i, \boldsymbol{y_{-i}}, \boldsymbol{r_{-i}})F(y_i - r_i) \quad (1)$$

Differentiating (1) with respect to y_i and r_i, respectively, yields the following first order conditions.[5]

$$\frac{\partial \Pi}{\partial y_i} = B'(y_i) - C'(Y) - \frac{\partial P(y_i, r_i, \boldsymbol{y_{-i}}, \boldsymbol{r_{-i}})}{\partial y_i} F(y_i - r_i) - P(y_i, r_i, \boldsymbol{y_{-i}}, \boldsymbol{r_{-i}})F'(y_i - r_i) = 0 \quad (2)$$

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

Equation (3) implies that the firm chooses the optimal report (degree of underreporting) to equate the additional cost of reporting (t) with the expected cost of underreporting one more unit. Note the equivalence of the last three terms in each of (2) and (3), recalling that $\frac{\partial P}{\partial y_i} = -\frac{\partial P}{\partial r_i}$. Thus (2) becomes

$$B'(y_i) - C'(Y) = t \quad (2')$$

The firm equates the marginal gain from output with cost of reporting. Assuming a symmetric equilibrium (y^*, r^*), the optimal choice of output (y^*) is defined by:

$$B'(y^*) - C'(ny^*) \equiv t \quad (4)$$

As shown in (4), the optimal level of output 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 output level is equivalent in the random and rank order tournament auditing schemes.

[5] The solution is the same whether we solve the firm's problem simultaneously or sequentially with the report being chosen in the second stage. This is because the expected fine depends only on the degree of underreporting and not *r* per se. Thus when choosing the report, the firm is choosing the amount of underreporting. This results from our specification of how the audit tournament operates and generates our independence result regarding output.

The intuition for this finding is as follows. At an interior solution, the firm will always choose its report such that the expected cost due to the regulatory mechanism of the marginal unreported unit is exactly *t*, which is the cost of reporting a unit. Therefore, the marginal cost of generating a unit of output is exactly *t* under any regulatory framework. This independence result occurs because our tournament depends on the degree of misreporting and would not occur under other assumptions. For example, if the regulator's estimate of the firm's output was unrelated to the firm's actual output and thus only affected by the report then this result would not hold.[6]

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

$$P(y^*, r_r^*)F'(y^* - r_r^*) \equiv t \tag{5}$$

Intuitively, the firm is comparing the reporting cost with the marginal expected fine. In this case, greater reporting reduces the fine size but has no impact on the inspection probability.

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

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

The additional (first) term reflects the fact 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. $P(y^*, r_r^*) = P(y^*, r_t^*)$. 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

[6] Alternative frameworks such as the one modeled by Evans et al. (2009) where the firm chooses a share of output to report and audits imperfectly reveal an output share could also lead to different results. We are grateful to an anonymous reviewer for providing this intuition and suggesting alternative frameworks.

higher under the tournament scheme than with random auditing, given identical output choices. This leads to our second main hypothesis.

Proposition 2: Reporting of output is greater under the rank order auditing tournament compared with random auditing.

Gilpatric et al. (2011) demonstrate that disclosure is higher with a rank order tournament or a more general relative audit mechanism than with random audits with a general error distribution but without the output choice. It is important to note however that we have demonstrated this result continues to hold even when output can be chosen.

The model discussed in this section focuses on the impact of varying the enforcement regime and does not explicitly model how social observability affects choices. Several approaches have been taken in the tax compliance literature to incorporate social norms in decision making (Hashimzade et al., 2013). One approach is to include a psychological cost from misreporting, where an individual's psychic cost parameter is affected by both their aversion to lying and the degree of social pressure (Gordon, 1989). Alternatively, others model a gain from truthful reporting, which increases with the degree of social conformity; that is, the gain is greatest when the proportion of truthful reporters is large (Myles and Naylor, 1996). Kim (2003) instead models the case where the psychic costs of misreporting occur only if you are exposed as a cheater. While the exact specifications differ, these models highlight how both individual attitudes and social pressure affect misreporting decisions. Further, social pressure is affected by both the social norm (the proportion of evaders in the population) and the magnitude of the psychic penalty for deviating from the social norm. The latter will be affected by the visibility of actions. We choose to examine these effects behaviorally and not formally model the impact of social observability because it is unclear which approach to adopt. Further, point predictions would be impossible to derive without making (strong) assumptions about the specific values for the psychic cost parameters.

4. Experimental Design

4.1 Decision Making

We conduct four treatments. In all treatments subjects interact in fixed groups of five and make decisions in 25 periods. Subjects participate in only one treatment. In all treatments and all periods, subjects make two decisions. The first is an output decision. This output generates a private benefit, $B(y_i)$, to the subjects, but imposes a cost on everyone in the group, $C(Y)$. Table 1 presents the functional forms and the parameters chosen for each of the variables in the experiment.[7] 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.[8] The maximum social benefit is 50 experimental dollars when the group total output is 100 units. At the individually-optimal output choice of 30 units (group total output 5×30=150), the net social benefit is 37.5 experimental dollars. 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 accurate revelation of actual output.

[7] These specific numerical predictions are based on risk neutrality. In the results section we discuss how the equilibrium changes with increasing risk aversion.

[8] Our design assumes that all agents derive similar private benefits and incur similar costs. In such a symmetric setting, it can be argued that in equilibrium everyone is inspected with similar likelihood, so outcomes in tournaments may converge to the random scheme. However in the previous section we show that reporting is higher under the rank order tournament compared to a random audit mechanism. In the experiment individuals exhibit some natural variation in preferences for creating negative externalities, for dishonesty aversion, and in costs of deviating from the perceived social norm. In future research heterogeneity may be an interesting extension to explore and this can have some relevant real world applications.

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 summarizing the fine for each unit of difference between the actual and reported output. The fine consists of two components: a convex 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 described earlier, the regulator observes output with some noise. As in the experiment the distribution of this noise term must be understood by subjects, we follow the standard practice in the literature of drawing ε_i from a uniform distribution because this is easily explained in the experiment instructions. Hence for simplicity, 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 $y_i = y_j = y^*$. As noted by Orrison et al. (2004), this implies that $\frac{\partial P(y_i, r_i, y_{-i}, r_{-i})}{\partial y_i} = -\frac{\partial P(y_i, r_i, y_{-i}, r_{-i})}{\partial r_i} = \frac{1}{2q}$. 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 $P(y^*, r_t^*) = m/n$ to determine the equilibrium amount of underreporting.

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.[9] The computer ranks the five members based on

[9] The particular choice of 17 for the maximum amount of noise makes the equilibrium level of misreporting substantially smaller in the Tournament than in the Random treatment, which is important so as to have a reasonable chance of identifying treatment differences in behavior. It also ensures that a pure strategy equilibrium exists.

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.[10] 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 computer screen. Among other things, displaying the photographs of participants captures the reputation effects that managers may 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. Photographs of group members were not shown in the Low Information treatments. The feedback screens for the High and Low Information treatments are presented in the instructions in the Appendix.

All sessions were conducted at the Vernon Smith Experimental Economics Laboratory at

[10] 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 output and enforcement concerning their peers, even though this information could be minimal in some cases. We also do not have treatments without auditing but with variations in observability since observability alone has been extensively explored in the literature, as described in Section 2. Further, such treatments would not be directly comparable to our current design which adds audit information feedback in the High information treatments.

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, 2015). Although some had participated in other economics experiments, all were inexperienced in the sense that they had never previously 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.

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. Five of these 25 periods were randomly drawn for payment at the conclusion of the session and experimental dollars were converted to U.S. dollars at a pre-announced 4-to-1 conversion rate. We used the terminology of output, 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 computerized quiz to examine and reinforce their comprehension and understanding of the instructions. The quiz was incentivized 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.

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. No subject raised any objections to their photograph being taken and used in the experiment.

After the 25 enforcement periods subjects participated in a brief risk assessment task before they learned which enforcement periods would be randomly drawn for payment. For this 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 behavior. 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. Subjects' total earnings averaged US$20.46 each, with an interquartile range of $13.25 to $26.75. Sessions usually lasted about 90 minutes on average, including the time taken for instructions and payment distribution.

4.2. Testable Hypotheses

The theoretical framework and the experimental design allow us to test the following hypotheses relating to output and reporting in the four treatments. First, Proposition 1 and 2 from our theoretical model (Section 3) 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: Under-reporting is greater in the Random enforcement treatment than the Tournament treatment.

Second, while we do not formally model the impact of social observability, our discussion of income tax compliance models incorporating social norms identified several channels through which social pressure might affect reporting decisions. The High information treatment could potentially operate via two of these channels. First, and most directly, revealing identities along with additional information about detected underreporting should increase the stigma from underreporting compared with the Low information treatment. Second, the additional information provides a signal about the proportion of evaders in the population, which might cause subjects to update their beliefs about the social norm and the cost of deviating from that norm. The first effect should decrease underreporting, while the direction of the second is unclear a priori. Similar considerations may also influence output choices. 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 underreport less. This leads to the following two conjectures.

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

Conjecture 2b: Underreporting will be lower in the High information treatment than the Low information treatment.

The effect of social observability could further 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 underreporting 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.

5. 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*. This allows us to include all types of misreporting, both over- and underreporting. A positive value indicates underreporting, a negative value overreporting, and zero indicates accurate reporting. The measure is dominated by underreporting, with 76% of the reports in this category, followed by 22% of reports being accurate. Only 2% of reports were overreports, which are clearly mistakes, and almost all were overreports of 10 or less. In Section 5.1 we include both positive and negative values of misreporting and consider misreporting on average, which is positive reflecting the dominance of underreporting. In Section 5.2 we focus only on underreporting.

5.1 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-value=0.05) but not in the High Information treatments (p-value=0.15). Pooling over information treatments reveals a statistically significant difference (p-value=0.02). Considering only output averaged across the last five periods yields similar conclusions.[11] Hypothesis 1a is therefore partially supported. There is no evidence to support Conjecture 1b.

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. As mentioned before, with our parameters the maximum social benefit is 50 experimental dollars and the net social benefit is 37.5 experimental dollars at the individually optimal output choice. Hence predicted efficiency is 37.5/50 = 75%. 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 exist 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 is different from 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-value=0.10) and Tournament (p-value=0.08) treatments. Looking only at the last five periods, the difference is

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

only significant for the random auditing treatment (p-value=0.01) but no longer in the tournament treatment (p-value=0.31).[12]

Thus, output tends to be modestly higher when using Random audits compared to the Tournament audit mechanism and with Random audits output is above the equilibrium prediction. Increased experience and learning leads output in the Tournament to tend toward the predicted value of 30.

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 clearly leads to a substantially lower level of misreporting than observed in the Random audit treatment. 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 treatment, but increases considerably with experience with Random auditing.

No significant differences exist across the information treatments in either the Random or the Tournament treatments when looking across all 25 periods or only the last five periods. Misreporting is significantly different between the Random and Tournament treatments whether disaggregated by information treatment (p-value=0.01 for Low; p-value=0.004 for High) or pooled across information treatments (p-value<0.0001). These differences persist in the last five periods. This evidence supports Hypothesis 2a but not Conjecture 2b.

With Random auditing, the predicted amount of misreporting is 21 while it is only 8.9 in the Tournament treatment. Pooling the information treatments, average misreporting over all periods

[12] These differences from the prediction cannot be explained by risk aversion because, as shown in (4), the output choice is unaffected by the enforcement parameters.

is significantly different than these predictions in both the Random (p-value=0.02) and Tournament treatments (p-value=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-value=0.24). The difference from 8.9 however remains in the late periods for the Tournament treatment (p-value=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 overreporting seen here is therefore unusual in the empirical literature.[13]

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.

5.2 Analysis of Underreporting Behavior

In the previous section we constructed a measure of the average amount of misreporting in the different treatments, where the measure included both over- and underreports. In this section we take a more detailed look at underreporting behavior by first investigating if the treatment differences are due to changes in the amount of underreporting or changes in the proportion of

[13] Risk aversion is an alternative explanation for the lower than predicted levels of misreporting. Numerical calculations using a constant relative risk aversion utility function reveal that the level of misreporting in the Random treatment is consistent with a moderate degree of risk aversion. For example, optimal misreporting declines to about 15 when the coefficient of relative risk aversion is 0.1. However risk aversion cannot explain the low levels of misreporting in the Tournament treatment. Equilibrium misreporting declines much more slowly with risk aversion in the strategic tournament, and with a coefficient of relative risk aversion of 0.5 it only falls to about 7.

subjects underreporting or both. We then investigate individual patterns in underreporting behavior. In this section we exclude the very small number of overreports from our analysis, thus all measures are of *under*reporting, rather than misreporting more generally.

Because the magnitude of underreporting 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 underreports in a particular period or reports accurately.

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-value<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.

As the bottom panel of Figure 3 shows, the proportion of underreporters remains roughly constant at a fairly high level in all treatments, at around 70-90% of subjects. This indicates 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.

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 dollars, which is less than one-third of the expected profits earned from optimal misreporting (about 24 experimental dollars). Most subjects (39 out of 60 in the Random treatment) maximally underreport in at least one period.

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 truthful or untruthful? Figure 5 shows the distribution of the number of periods (out of 25) of truthful reporting (i.e. report=output) for each of the 120 subjects, separately for the two audit treatments. The figure shows that the majority of subjects misreported in most periods. In the Random treatment, 15% of subjects misreported in every period (i.e. number of truthful periods=0) compared to 20% of subjects in the Tournament treatment. More than half of the subjects were truthful in only 3 periods or less (62% in Random and 53% in Tournament). Pooled across treatments 75 of the 120 subjects misreported in at least 80 percent of all periods.

No one reported truthfully in every single period; two people were truthful in 24 periods and one in 21 periods.

5.3 Individual Motivations and Characteristics

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. Subjects self-select into the experiment and are not a representative sample, so their responses cannot be interpreted as revealing motivations broadly held by any general population. Nevertheless, the responses do help us interpret the behavior they exhibited in 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. This may be one reason why the social observability of compliance choices did not affect behavior in the experiment.

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-value<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-value<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.

Appendix A reports panel regressions that examine individual level heterogeneity in output and reporting choices. In addition to the experimental treatment variables and a time trend, the random effects regressions include variables that explore the dynamics in the reporting and output behavior. We create lagged variables that capture the average amount reported by other group members in the previous period, information which is provided to subjects in all treatments on the feedback screen, and a lag variable that records if a subject was inspected in the previous period. In the High information treatments, subjects are also informed about the detected underreporting of inspected subjects. For this treatment we also construct a measure of revealed underreporting by other subjects in the previous period. Demographic variables and questionnaire responses are also included in the regressions.

Across all specifications, the treatment effects described earlier remain robust. The Tournament audit leads to significantly lower output and lower underreporting, by all measures, while High information has no significant impact. This superior performance of the Tournament for reporting is consistent with Proposition 2 and Hypothesis 2a. This is intuitive since greater misreporting increases the inspection likelihood only in the Tournament, and fines (conditional on the amount of underreporting) are equivalent in the two mechanisms.

With regards to the dynamics of output and reporting choices, we find that subjects learn to choose higher output levels and misreport more over time in the Random mechanism while this is not the case in the Tournament mechanism.[14] We also find that when subjects observe higher average output reported in the previous period by others in their group, this led to higher output in the Tournament treatment but not in the Random treatment. Being inspected in the previous period increases the amount of output in the Random treatment. It also increases the amount and percentage of underreporting in the Random treatment but not in the Tournament treatment. In the Random treatment, subjects behave in a manner consistent with the gambler's fallacy of misunderstanding serially uncorrelated random events. In particular, after being inspected in the previous period they increase both output and underreporting, which suggests that they anticipate a decreased likelihood of being inspected in the subsequent period.[15] The estimates do not provide any evidence that revealed underreporting of other subjects in the previous period affects output or misreporting.

Amongst the individual level controls, subjects who reported being extremely motivated by social stigma concerns chose a significantly smaller output amount. Males and less risk averse subjects (as measured by preferences elicited by the incentivized risk task) misreport more, while subjects with high self-reported grades misreport less. The higher misreporting observed for less risk averse subjects is consistent with the inverse equilibrium relationship between these variables mentioned earlier in footnote 13. Further specifications that explore the effect of personal characteristics provide some evidence that their influence is mitigated in the

[14] Note the significant time trend in the top line of the tables (Tables A1-A3) is typically matched with a roughly equivalent estimate of opposite sign for the Tournament*trend interaction, indicating these time trends are only present in the Random mechanism.

[15] In an alternate specification examining the likelihood of maximal misreporting (Column 5, Table A3 in Appendix A), subjects were observed to more frequently misreport maximally if they were inspected in the previous period, but only in the Random treatments.

Tournament treatments. This suggests that subjects likely focus on the economic incentives induced by the mechanism more in this treatment than with Random audits.[16]

6. Conclusion

Designing appropriate audit mechanisms that improve compliance outcomes cost-effectively is critical for improving regulatory efficiency. In this paper we study and compare two audit mechanisms. While one is based on simple random audits the other has its roots in competition and tournaments and can be used to target firms and individuals suspected of being noncompliant. We also vary orthogonally the social observability of actions, and focus on two main performance measures - output and reporting.

Although optimal output levels are theoretically independent of the audit schemes, output in the tournament treatment of the experiment is lower (i.e., more socially optimal) than in the random audit treatment. While many subjects might correctly recognize that their audit probability depends on the output-report gap, some could mistakenly alter output rather than focusing entirely on reporting. These output treatment differences are modest, and output is generally close to noncooperative Nash equilibrium levels; efficiency, however, is substantially higher in the tournament treatment. Thus using competition in enforcement can have benefits that extend beyond simply encouraging more truthful reporting. 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.

[16] Detailed results from these specifications that interact individual characteristics with the audit treatments are available from the authors.

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 impact on output or reporting in either of the treatments. This is contrary to a recent 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 an 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 which suggests that misreporting is generally acceptable in this context. Thus being exposed as an untruthful 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. Further, for some individuals being in a laboratory setting that focuses on misreporting and penalties could have triggered experimenter demand effects, and this in conjunction with observing misreports frequently, may have led them to misreport more often.[17] These factors could have contributed to the observed ineffectiveness of social observability in our study, and the conclusion that information disclosure is neither beneficial nor detrimental to formal regulations.

[17] The impact of observability could also depend on the context. Our experimental task specified costs and penalties, thus creating a setting that subjects would particularly connect to auditing. It is still possible that some subjects regarded the decision making in this task as a socially distant, lottery or game- like task and hence were less influenced by social observability. A recent paper by Bartling et al (2015) reiterates the importance of context in decision making and shows that subjects interacting in market situations exhibit less social concern than those in comparable individual choice contexts. The impact of observability could also depend on whether social information is disclosed to peers (such as other firms) or to other agents (such as consumers) who may evaluate performance on different measures. Future research could aim to disentangle these effects.

These findings should be viewed with caution since they arise from a simple laboratory experiment, but they suggest that social information about output and reporting choices can have subtle effects. If similar effects also exist in practice, social observability measures such as public disclosure of toxic chemical releases (e.g., US EPA, 2014b) may have some 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 behavioral differences between them increase substantially over time.

As policymakers are starting to incorporate non-random audit schemes in their compliance toolkit, it is important to understand the underlying motivations for compliance induced by these new schemes. For example, federal tax authorities in the U.S. are much more likely to audit individuals who report 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.[18] In other regulatory areas, it may be difficult

[18] 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.

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.

Acknowledgement: The authors thank many colleagues, the editor, and two anonymous referees for helpful comments, Fatemeh Momeni and Huanren Zhang for research assistance, and seminar and conference participants at ANZWEE 2014, ESA North American Meeting 2014, the University of Kansas, Deakin University, University of Lyon (GATE), ASFEE Conference 2015 (Paris), Monash Sustainability Workshop 2015 (Prato) and the ESA World Meetings 2015. Funding from the Cooperative Research Centre for Water Sensitive Cities (CRC grant number 20110044) is gratefully acknowledged. We of course retain responsibility for any errors or omissions.

References

Andreoni, J., and R. Petrie. 2004. “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.

Bartling, B., R. Weber and L. Yao. 2015. “Do Markets Erode Social Responsibility.” *Quarterly Journal of Economics*, 1-48.

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.

Coricelli, G., E. Rusconi and M. C. Villeval. 2014. “Tax Evasion and Emotions: An Empirical Test of Re-Integrative Shaming Theory.” *Journal of Economic Psychology*, 40, 49-61.

Dickinson, D., E. G. Dutcher and C. Rodet. 2015. “Observed Punishment Spillover Effects: A Laboratory Investigation of Behavior in a Social Dilemma.” *Experimental Economics*, 18, 136-153.

Dufwenberg, M., and A. Muren. 2006. “Generosity, Anonymity, Gender.” *Journal of Economic Behavior and Organization*, 61, 42-49.

Evans, M. F., S. M. Gilpatric and L. Liu. 2009. “Regulation with Direct Benefits of Information Disclosure and Imperfect Monitoring.” *Journal of Environmental Economics and Management*, 57, 284-292.

Fischbacher, Urs. 2007. “z-Tree: Zurich Toolbox for Ready-Made Economics Experiments.” *Experimental Economics*, 10, 171-178.

Gilpatric, S., C. Vossler and L. Liu. 2015. “Using Competition to Stimulate Regulatory Compliance: a Tournament-Based Dynamic Targeting Mechanism.” *Journal of Economic Behavior and Organization*, 119, 182-196.

Gilpatric, S., C. Vossler and M. McKee. 2011. “Regulatory Enforcement with Competitive Endogenous Audit Mechanisms.” *RAND Journal of Economics*, 42(2), 292-312.

Gordon, J. P. P. 1989. “Individual Morality and Reputation Costs as Deterrents to Tax Evasion.” *European Economic Review*, 33(4), 797-805.

Greiner, Ben. 2015. “Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE.” *Journal of the Economic Science Association*, 1, 114-125.

Hashimzade, N., G. D. Myles and B. Tran-Nam. 2013. “Applications of Behavioral Economics to Tax Evasion.” *Journal of Economic Surveys*, 27(5), 941-977.

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.

Kim, Y. 2003. “Income Distribution and Equilibrium Multiplicity in a Stigma-Based Model of Tax Evasion.” *Journal of Public Economics*, 87(7–8), 1591-1616.

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.

Myles, G. D., and R.A. Naylor. 1996. “A Model of Tax Evasion with Group Conformity and Social Customs.” *European Journal of Political Economy*, 12(1), 49-66

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. 2015. “Firms’ Emissions and Self-Reporting under Competitive Audit Mechanisms.” *Environmental and Resource Economics*, 62, 949-978.

Orrison, A., A. Schotter and K. Weigelt. 2004. “Multiperson Tournaments: An Experimental Examination,” *Management Science*, 50, 268-279.

Perez-Truglia, R., and U. Troiano. 2015. “Tax Debt Enforcement: Theory and Evidence from a Field Experiment in the United States” Working Paper.

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. Sheremeta. 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. 2015. “A Note on Stochastic Public Revelation and Voluntary Contributions to Public Goods”, *Economics Letters*, 126, 144-146.

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. 2014a. *Fiscal Year 2014-2016 EPA Strategic Plan.* Available at http://www2.epa.gov/planandbudget/strategicplan (accessed 21 October 2015).

U.S. Environmental Protection Agency. 2014b. *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
y_i	Output of firm i	Output ranges from 0 to 40 units.
$B(y_i) = y_i$	Benefit of firm i from Output chosen	Each unit of output generates $1 for the firm
$C(Y)= Y^2/1000$	Cost of sum of output chosen by all firms	Depends on output chosen by the 5 firms
t	Per unit reporting cost	0.7
$P(y^*, r_r^*) = P(y^*, r_t^*)$	Probability of Inspection	0.4
$[-q, +q]$	For Tournament Audit: The noise added to their actual output.	[-17, +17]
$F(y_i - r_i) = (y_i - r_i)^2/40 + t(y_i - r_i)$;	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**
Equilibrium choice of output (y*)	30	30
Equilibrium choice of reporting (r*)	9	21.1
Equilibrium level of misreporting (y*-r*)	21	8.9
Socially Optimal choice of output	20	20

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

Table 3: 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**
Random	*Low*	31.1 (3.8)	32.5 (3.1)	14.8 (7.1)	17.4 (6.9)
Random	*High*	31.6 (2.5)	34.8 (3.1)	15.1 (3.0)	20.8 (3.7)
Tournament	*Low*	28.4 (1.7)	28.7 (3.6)	4.0 (1.9)	3.8 (2.2)
Tournament	*High*	28.5 (3.8)	28.1 (6.5)	4.5 (2.2)	5.1 (2.4)

Note: Means and standard deviations calculated over the statistically independent groups of five subjects.

Table 4: Average Degree of Underreporting and Percentage of Underreporters by Treatment (Standard Deviation)

Treatment		**Average % Underreport**		**Average Percentage of Underreporters**	
Audit	**Information**	**All Periods**	**Last Five Periods**	**All Periods**	**Last Five Periods**
Random	*Low*	43.2 (17.6)	48.7 (17.8)	81.3 (12.4)	82.6 (14.4)
Random	*High*	44.1 (6.7)	55.2 (8.8)	79.9 (9.5)	86.7 (11.8)
Tournament	*Low*	15.5 (5.6)	15.1 (7.1)	76.3 (12.5)	83.5 (12.6)
Tournament	*High*	17.8 (10.8)	19.9 (12.5)	71.4 (22.3)	75.1 (21.3)

Note: Means and standard deviations calculated over the statistically independent groups of five subjects.

Table 5: Motivations for Obeying Laws and Misreporting in the Experiment

% of responses in each category	*Not at all important*	*Slightly important*	*Somewhat Important*	*Important*	*Extremely Important*
General Law Obedience (Q14)					
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
Accurate reporting in experiment (Q18)					
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* <u>*personal*</u> *motivations for your decision to obey or not obey a law. Question 18:* (For experimental choices) *Please indicate how you would rate the importance of the following motivations for your decision to accurately report or not accurately report your output.*

Figure 1: Average Output by Treatment

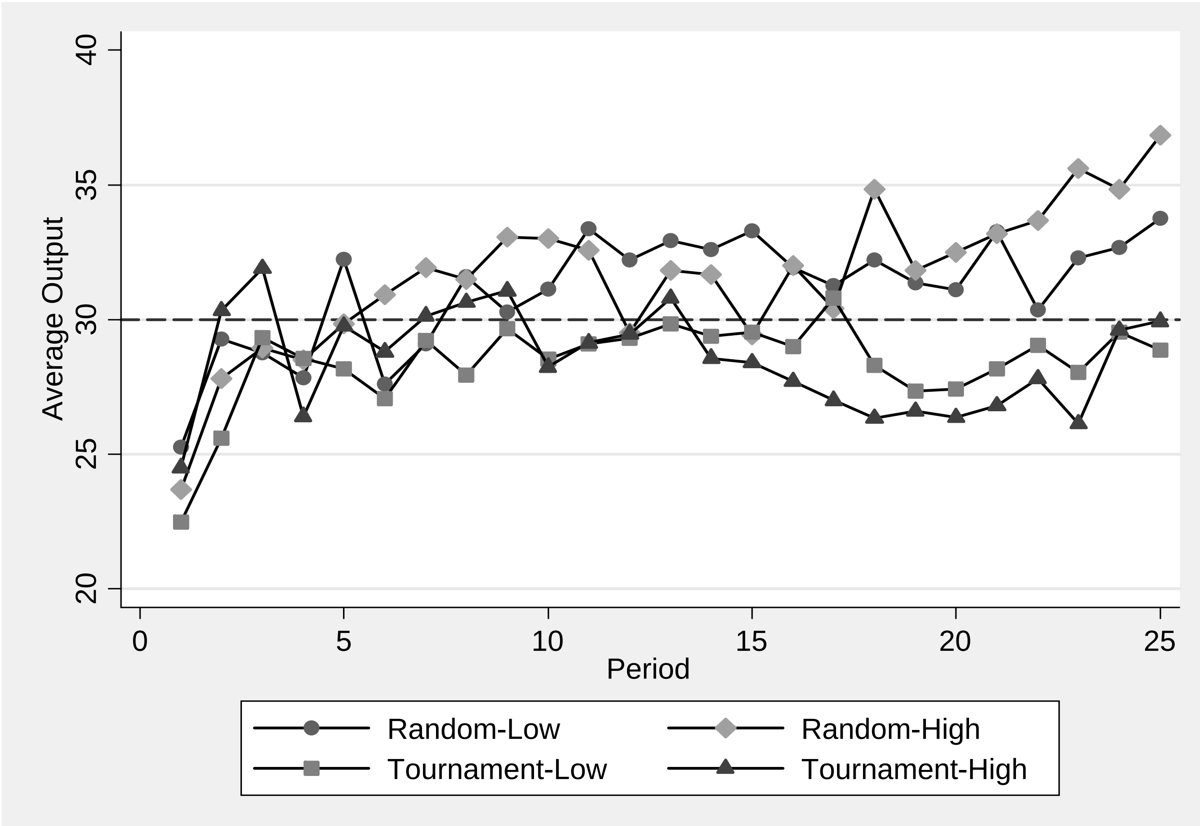


Figure 2: Average Misreport by Treatment

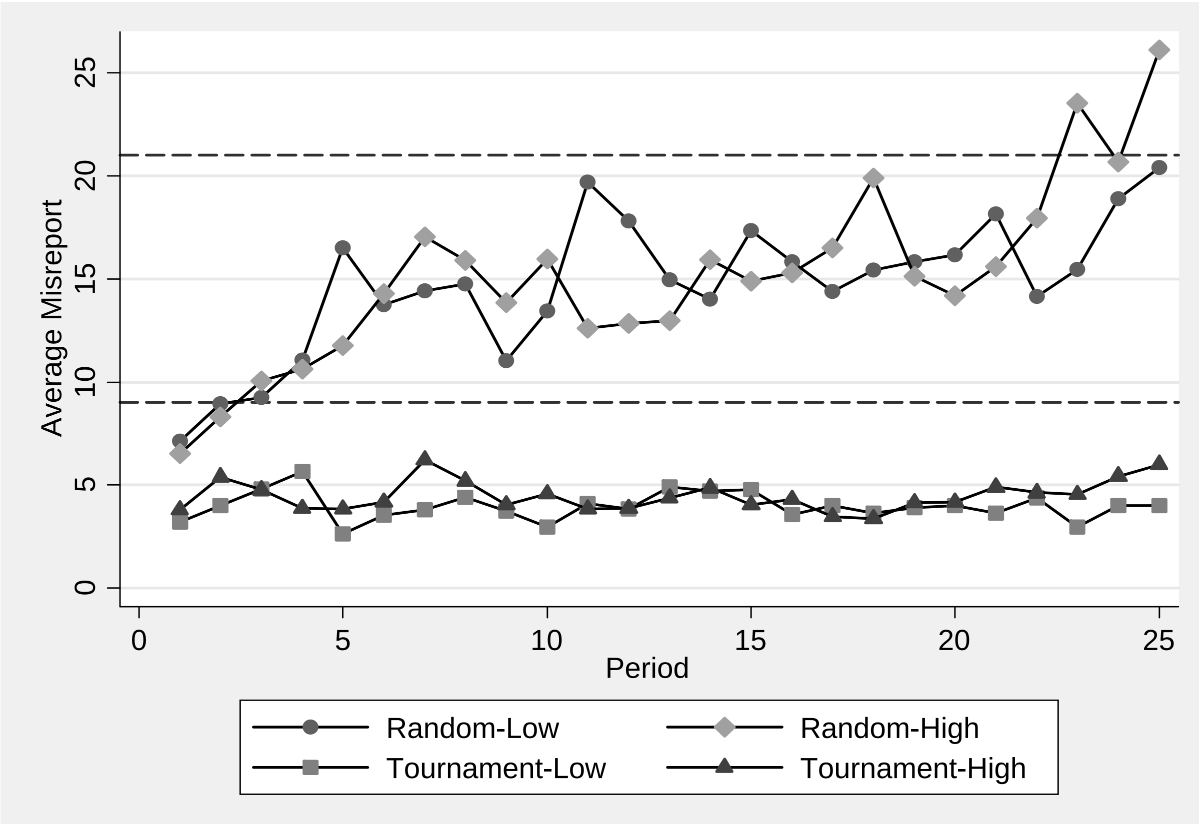
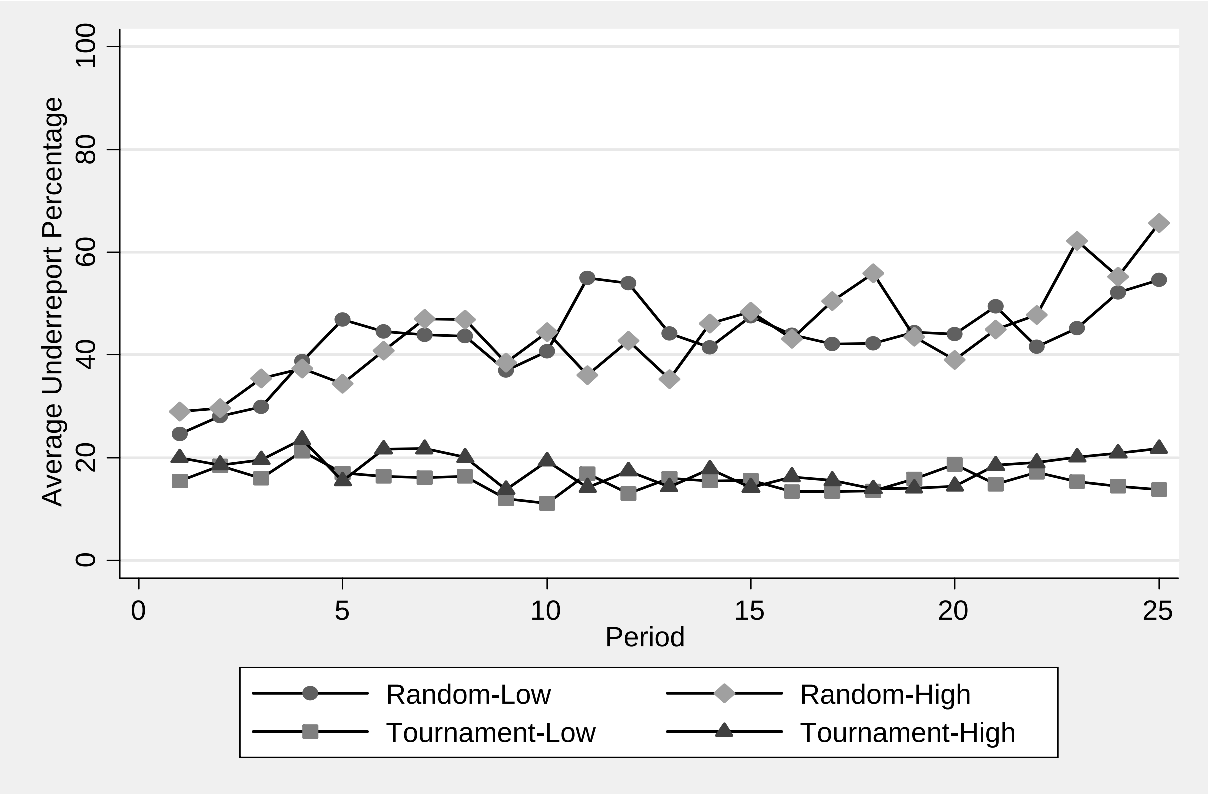


Figure 3: Average Degree of Underreporting and Percentage of Underreporters by Treatment



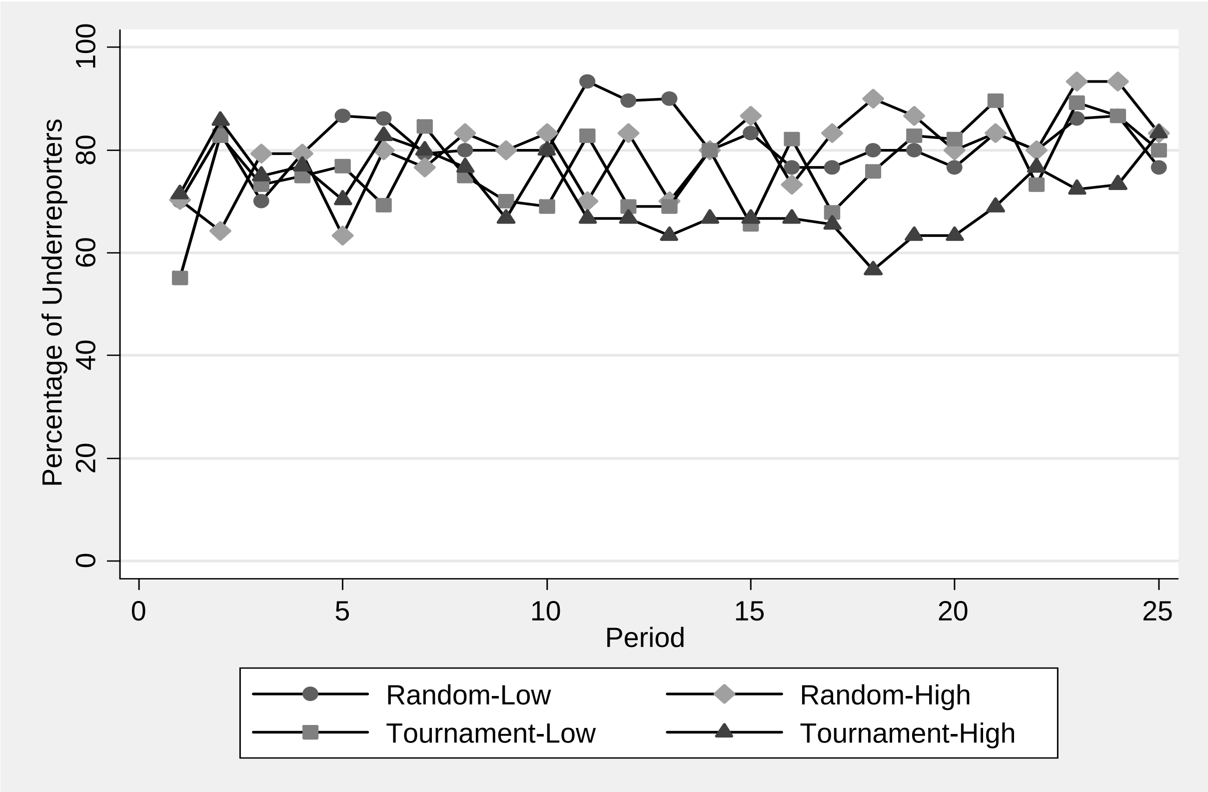


Figure 4: Distribution of Degree of Underreporting by Audit Treatment

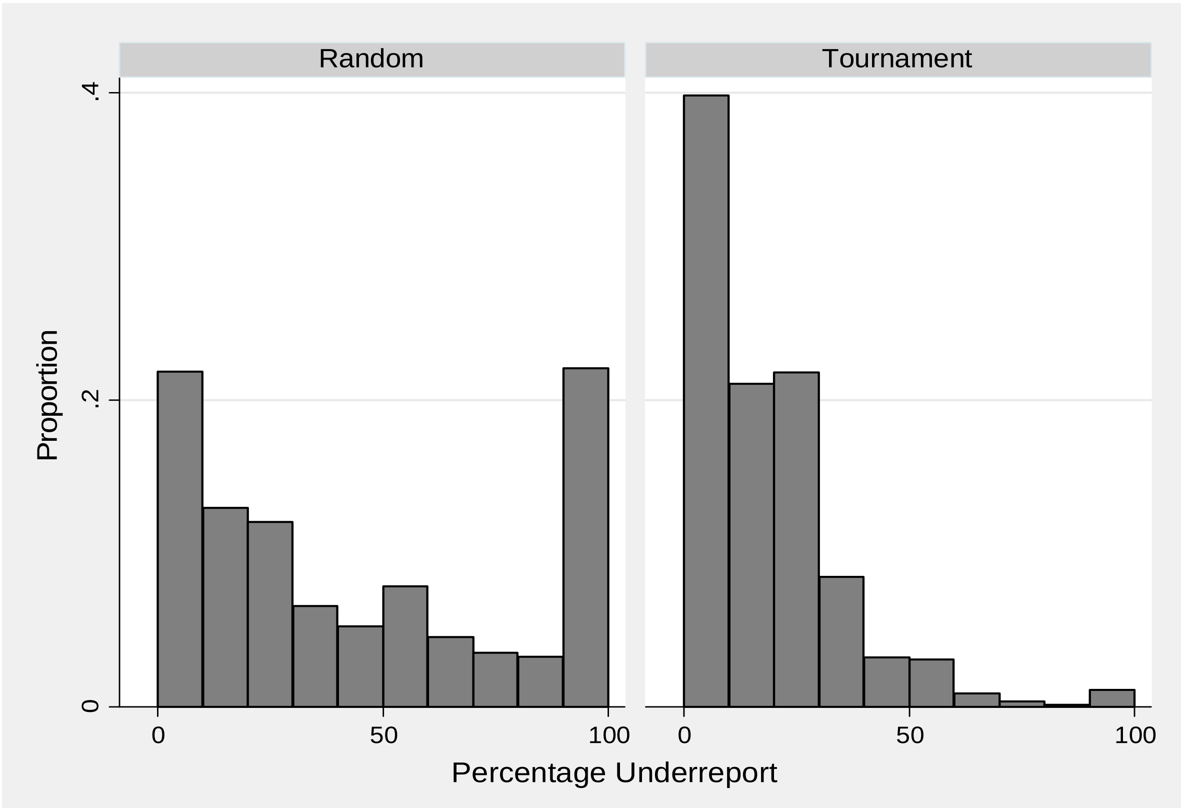
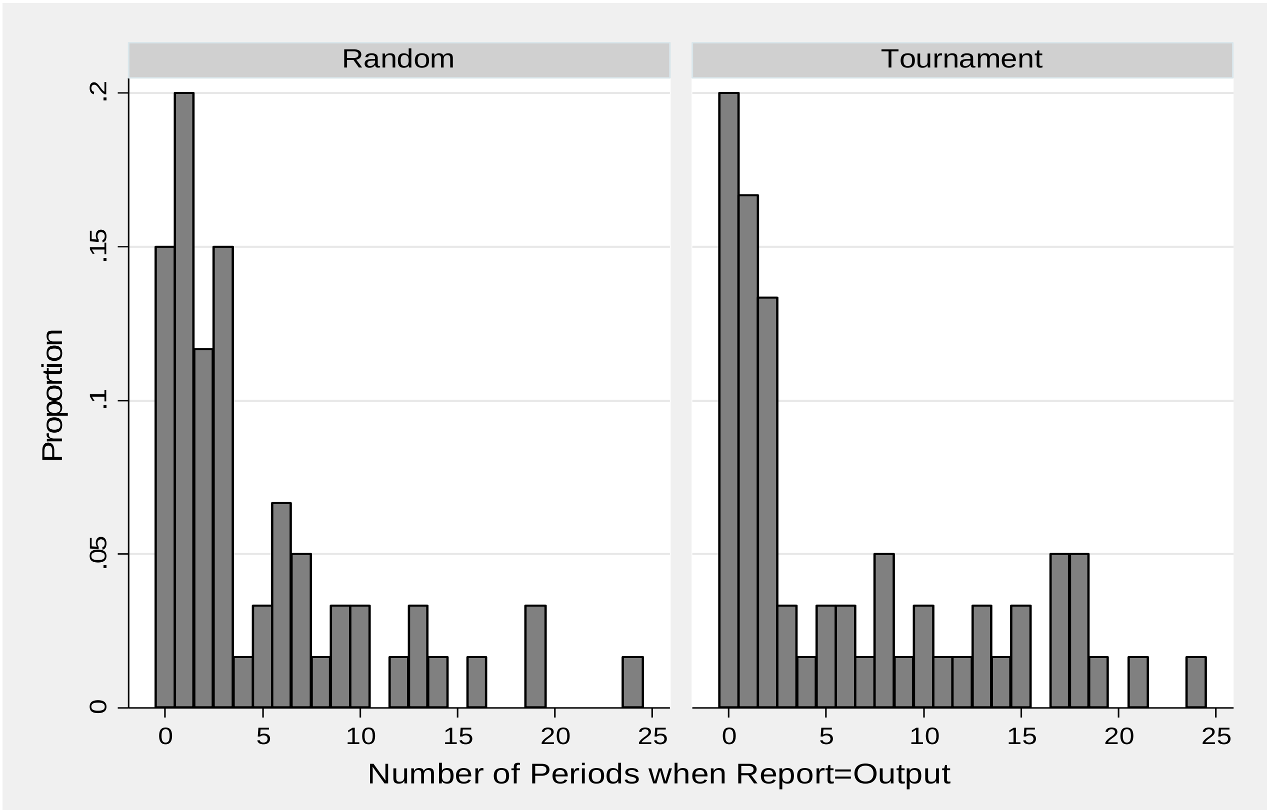


Figure 5: Individual Subject Frequencies of Number of Periods of Truthful Reporting



Appendix A: Panel Regressions (including Subject Characteristics) of Output and Reporting Choices

In this section we present results from panel regressions that employ random effects at the individual level and clustering at the group level. Estimates using output, misreport, or %underreport as the dependent variable are reported in Tables A1, A2, and A3 respectively. In each regression we include the following experimental variables: indicators for the two treatments (High information and Tournament), separately and interacted with each other, a trend variable (1/period), and the interaction between Tournament and trend, as explanatory variables. We also include various demographic control variables such as gender (*male*), place of birth (*US born*) and having a self-reported 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*).[19] They also include indicators derived from the questionnaire responses 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 economics experiments.

The different specifications include various lagged measures to study the dynamics of behavior. In particular, *LagReport* is the average amount reported by other subjects in the previous period; *LagInspected* is a dummy variable equal to one if a subject was inspected in the previous period; and *LagUnderReport* is the average amount of underreporting observed for the other subjects who were inspected in the previous period. These are discussed in detail in section 5.3.

[19] 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-value=0.001, Mann-Whitney test). A quarter (26%) of participants state they have no religion, while 44% identify as Christian.

Results from all of these models confirm our earlier findings regarding treatment differences. The indicator variable for High information is never significant either separately or when interacted with Tournament. On the other hand, the Tournament variable is always highly significant, indicating that this enforcement mechanism leads to reduced output, and a lower amount of misreporting and 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 audit treatments.

Finally, note that the demographic and additional control variables have little explanatory power for output choices, although interestingly subjects who reported being extremely motivated by social stigma concerns chose a significantly smaller output amount. Males, less risk averse subjects 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. The lower misreporting for more risk averse subjects is consistent with the inverse equilibrium relationship between these variables mentioned in section 5.1. Subjects who indicate they have "no religion" or are "Christian" misreport more than those from other religions. Motivational factors expressed in the post-experiment questionnaire were insignificant determinants of misreporting.[20]

[20] 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 important). As these variables were never significant, either individually or jointly, we do not report them here.

Table A1: Panel Regressions of Output

Dependent Variable:	(1) output	(2) output	(3) output	(4) output
trend (1/Period)	-9.04***	-11.92***	-11.74***	-15.40***
	(1.77)	(3.29)	(3.22)	(4.12)
High Information	-0.10	-0.01	-0.07	
	(1.66)	(1.75)	(1.73)	
Tournament	-4.01**	-10.03***	-4.17**	-6.80***
	(1.95)	(2.76)	(2.12)	(2.41)
High Information * Tournament	-0.20	-0.30	-0.34	
	(2.38)	(2.28)	(2.48)	
Tournament * trend	5.19**	14.73***	13.19**	22.52***
	(2.64)	(4.46)	(5.19)	(7.17)
Male	0.02	0.13	-0.06	-1.80*
	(0.83)	(0.82)	(0.86)	(1.10)
US born	0.39	0.39	0.44	0.90
	(1.39)	(1.32)	(1.41)	(2.35)
high GPA	0.18	-0.04	0.16	-0.33
	(1.58)	(1.62)	(1.61)	(2.31)
experienced subject	-2.27	-2.37	-2.44	-3.17**
	(1.47)	(1.46)	(1.49)	(1.40)
no religion	1.08	0.82	1.18	-0.07
	(1.04)	(1.10)	(1.06)	(1.57)
Christian	-0.03	-0.01	-0.08	-1.02
	(1.16)	(1.21)	(1.21)	(1.70)
Amount Invested in the Risk Task	0.36	0.38	0.37	0.57
	(0.26)	(0.26)	(0.26)	(0.41)
stigma extremely important	-4.80**	-5.16**	-4.93**	-4.38
experimental motivation	(2.24)	(2.36)	(2.29)	(3.11)
penalty extremely important	-1.21	-1.42	-1.33	-1.51
experimental motivation	(1.29)	(1.38)	(1.30)	(2.33)
LagReport		0.03		
		(0.05)		
LagReport * Tournament		0.20**		
		(0.09)		
LagInspected			1.67***	
			(0.55)	
LagInspected * Tournament			-1.58**	
			(0.73)	
LagUnderReport				-0.02
				(0.02)
LagUnderReport * Tournament				0.06
				(0.08)
Constant	33.68***	33.67***	33.37***	36.66***
	(2.14)	(2.25)	(2.19)	(1.58)
Observations	3000	2880	2880	1380
Subjects	120	120	120	60
Wald test of model (p-value)	0.0000	0.0000	0.0000	.

Clustered standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Panel Regressions of Misreporting

Dependent Variable:	(1) misreport	(2) misreport	(3) misreport	(4) misreport
trend (1/Period)	-12.51^{***}	-20.13^{***}	-21.06^{***}	-9.83^{***}
	(2.38)	(4.03)	(4.20)	(2.76)
High Information	0.55	0.57	0.55	
	(2.66)	(2.56)	(2.79)	
Tournament	-12.28^{***}	-14.70^{***}	-10.60^{***}	-6.09^{***}
	(2.91)	(3.54)	(2.97)	(1.36)
High Information * Tournament	-0.88	-0.92	-0.92	
	(3.05)	(2.95)	(3.17)	
Tournament * trend	12.61^{***}	22.06^{***}	23.16^{***}	12.65^{***}
	(2.46)	(4.15)	(4.28)	(2.96)
Male	1.80^{*}	1.83^{*}	1.81^{*}	0.28^{**}
	(0.97)	(0.97)	(1.05)	(0.51)
US born	-1.41	-1.22	-1.31	-1.84^{***}
	(1.31)	(1.30)	(1.31)	(0.53)
high GPA	-3.01^{**}	-3.04^{**}	-3.12^{**}	-2.92^{***}
	(1.44)	(1.515)	(1.47)	(0.72)
experienced subject	3.69^{**}	3.62^{**}	3.49^{**}	2.55^{**}
	(1.55)	(1.58)	(1.61)	(1.29)
no religion	2.58^{***}	2.76^{***}	2.85^{***}	1.15^{*}
	(0.84)	(0.86)	(0.86)	(0.63)
Christian	2.12^{**}	2.02^{**}	2.25^{**}	2.32^{**}
	(0.91)	(0.88)	(0.92)	(1.16)
Amount Invested in the Risk Task	0.97^{***}	0.89^{***}	0.90^{***}	0.40
	(0.24)	(0.24)	(0.23)	(0.32)
stigma extremely important	-2.15	-1.97	-1.99	-3.01^{***}
experimental motivation	(2.37)	(2.35)	(2.54)	(0.82)
penalty extremely important	-1.11	-1.18	-1.38	-0.68
experimental motivation	(1.26)	(1.23)	(1.28)	(1.00)
LagReport		-0.13		
		(0.11)		
LagReport * Tournament		0.10		
		(0.11)		
LagInspected			6.13^{***}	
			(0.79)	
LagInspected * Tournament			-6.89^{***}	
			(0.81)	
LagUnderReport				-0.02
				(0.06)
LagUnderReport * Tournament				-0.01
				(0.07)
Constant	16.38^{***}	19.11^{***}	14.68^{***}	8.66^{***}
	(3.10)	(3.81)	(3.20)	(1.34)
Observations	3000	2880	2880	1380
Subjects	120	120	120	60
Wald test of model (p-value)	0.0000	0.0000	0.0000	.

Clustered standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Panel Regressions of % Under-Reporting

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	%underreport	%underreport	%underreport	%underreport	maxunder[a]
trend (1/Period)	-26.29***	-42.64***	-44.71***	-45.70***	-3.51***
	(6.69)	(11.10)	(11.34)	(16.59)	(0.73)
High Information	2.47	2.20	2.18		0.01
	(6.44)	(6.15)	(6.74)		(0.54)
Tournament	-29.26***	-32.23***	-24.90***	-36.82***	-2.86**
	(7.28)	(8.87)	(7.38)	(6.54)	(1.41)
High Information * Tournament	-2.95	-2.83	-2.82		0.73
	(7.98)	(7.60)	(8.26)		(1.09)
Tournament * trend	30.49***	52.17***	55.54***	56.04***	4.41***
	(7.17)	(12.04)	(12.32)	(16.52)	(0.81)
Male	6.65**	6.75**	6.80**	9.68***	0.57**
	(2.79)	(2.81)	(3.01)	(3.22)	(0.28)
US born	-5.57	-4.93	-5.17	-9.98**	-0.64***
	(3.58)	(3.54)	(3.60)	(4.22)	(0.21)
high GPA	-7.33*	-7.33*	-7.68**	-14.17***	-0.43
	(3.84)	(3.99)	(3.89)	(4.75)	(0.34)
experienced subject	19.05***	18.82***	18.49***	23.56**	0.77**
	(4.79)	(4.86)	(4.88)	(9.51)	(0.34)
no religion	7.37***	7.64***	7.69***	8.61**	0.59***
	(2.86)	(2.85)	(2.84)	(4.05)	(0.23)
Christian	6.33***	5.81**	6.41***	7.68	0.56*
	(2.35)	(2.33)	(2.35)	(4.78)	(0.29)
Amount Invested in Risk Task	1.92***	1.74***	1.78***	2.51***	0.17***
	(0.53)	(0.54)	(0.52)	(0.78)	(0.03)
stigma extremely important	-4.74	-3.85	-4.03	-17.93***	0.30
experimental motivation	(6.03)	(5.95)	(6.27)	(3.96)	(0.31)
penalty extremely important	-3.72	-3.83	-4.37	-5.02	-0.01
experimental motivation	(3.08)	(3.03)	(3.14)	(3.43)	(0.34)
LagReport		-0.32			
		(0.27)			
LagReport * Tournament		0.13			
		(0.29)			
LagInspected			15.23***		0.76***
			(2.27)		(0.10)
LagInspected * Tournament			-17.62***		-0.87***
			(2.39)		(0.28)
LagUnderReport				0.01	
				(0.21)	
LagUnderReport * Tournament				0.04	
				(0.25)	
Constant	43.32***	49.97***	39.028***	50.74***	-1.47*
	(8.01)	(9.53)	(8.12)	(7.14)	(0.76)
Observations	2937	2823	2823	1361	2823
Subjects	120	120	120	60	120
Wald test of model (p-value)	0.0000	0.0000	0.0000	.	0.0000

Clustered standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a] Probit regression of the likelihood of misreporting to the maximum degree possible.

Appendix B: Experiment Instructions (Tournament-High Information Treatment)

Introduction

This experiment is a study of group and individual decision making. The amount of money you earn depends partly on the decisions that you make and thus you should read the instructions carefully. The money you earn will be paid privately to you, in cash, at the end of the experiment. A research foundation has provided the funds for this study. Please put away your cell phones and other distracting devices for the duration of the experiment.

In this experiment, you will be in a group consisting of five people, you and four others. The other people in your group are sitting in this room. You will make decisions privately, that is, without consulting other group members. Please do not attempt to communicate with other participants in the room during the experiment. If you have a question as we read through the instructions or any time during the experiment, please raise your hand and an experimenter will come by to answer it. At the end of these instructions you will take a computerized quiz and earn $0.50 for each correct answer you provide.

The experiment is divided into 25 decision "periods." You will be paid based on your decision in five randomly chosen periods. Each decision you make is therefore important because it has a chance of determining the amount of money you earn. Your group members remain the same from one period to another.

Your earnings in the experiment are denominated in experimental dollars, which will be exchanged at a rate of 4 experimental dollars = 1 U.S. dollar at the end of the experiment.

Overview

You are given an initial earnings of 20 experimental dollars every period.

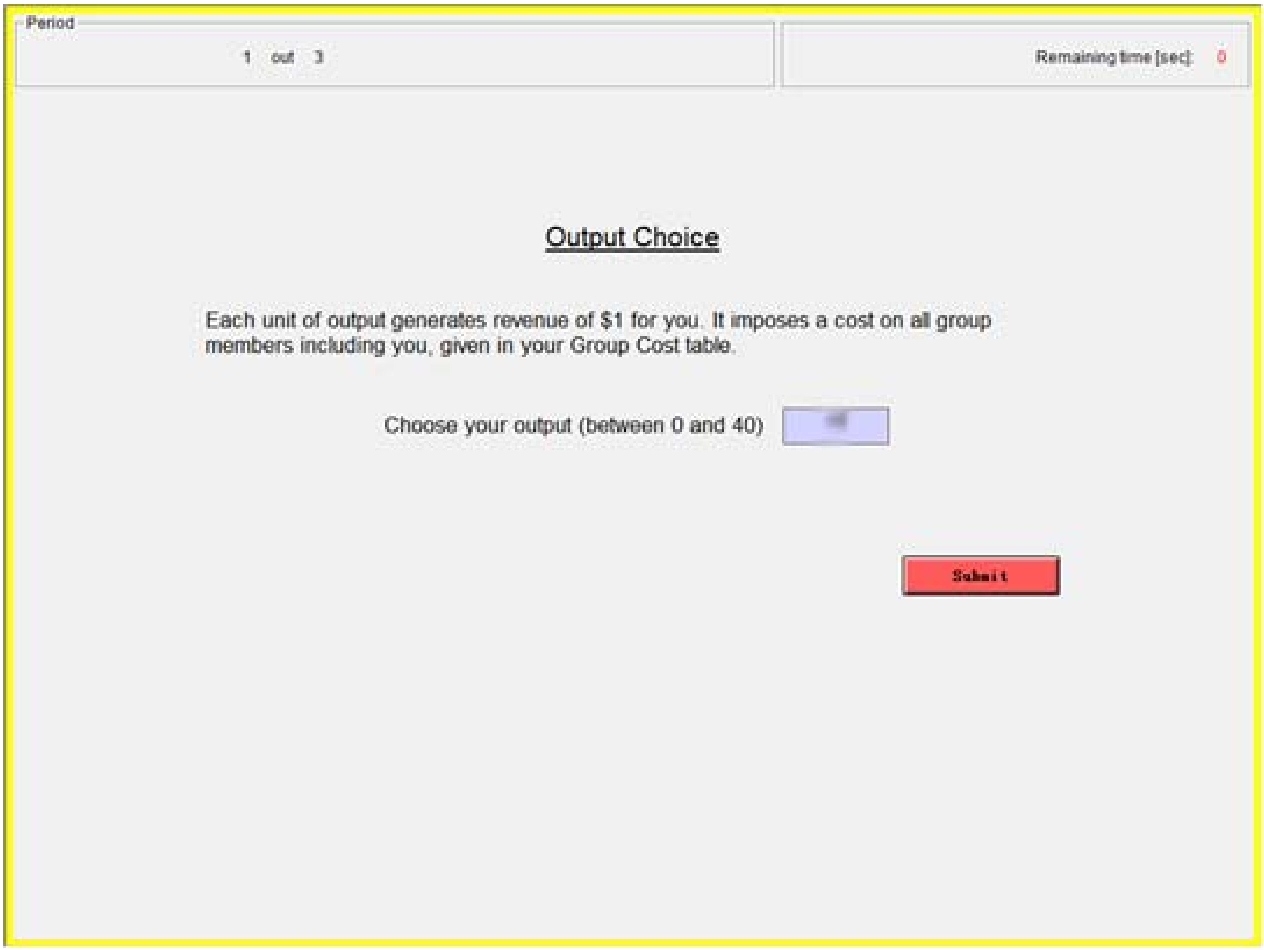
Each decision period, you need to decide how much output to produce and then decide how much to report. You pay a cost based on how much output you report.

The difference between the amount of output you produce and how much you report affects the chance that you will be inspected. If you are inspected, and found to have reported less than your actual output you may face additional costs.

Each decision period is independent from the others, in the sense that your earnings in one period depend only on your decision- and the decisions of others- in that particular period.

Your decisions

In each period you will make two decisions. Your first decision (and those of others in the group) is to choose an amount of output to produce and to enter it into the computer. This must be between 0 and 40. Your decision will be entered on a screen like the one on the next page.



Each unit of output generates a revenue of $1 for you, but also imposes a cost on all group members including yourself. This cost takes the following form:
Group cost = (Sum of output chosen by all 5 participants)^2/1000.

This Group cost depends on the total amount of output chosen by everyone in the group and is the same for everyone in the group. Table 1 provides information about the group cost for the total amount of output chosen by the group.

Hence your earnings from this output decision in each period is the difference between your private revenue and the group cost from choosing this output:
Earnings = Private revenue - Group cost
The higher the output you choose the higher will be your revenue and the higher will be the group cost. For example, if the sum of the output chosen by all group members is 110, then the group cost incurred by each group member is $12.1 (as seen in the table below). If you increase your output by 10, you increase your revenue by $1×10=$10. But you also increase your group cost (and everyone else's group cost in your group) from $12.1 to $14.4 because your output increase raises the group sum output from 110 to 120.

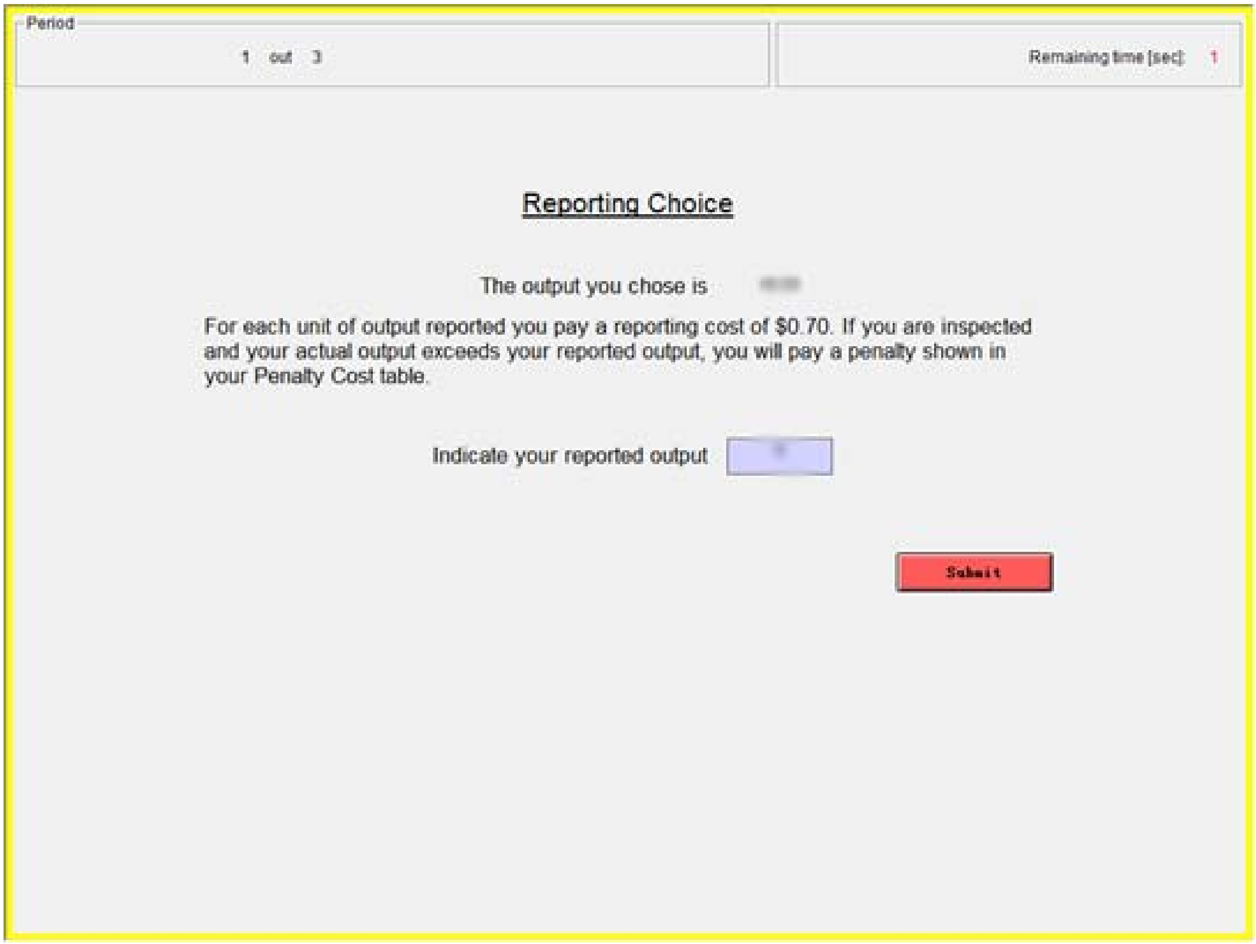
Table 1: Group Cost

Sum of Output (all 5 Participants)	Group Cost	Sum of Output (all 5 Participants)	Group Cost
0	0	105	11.025
5	0.025	110	12.1
10	0.1	115	13.225
15	0.225	120	14.4
20	0.4	125	15.625
25	0.625	130	16.9
30	0.9	135	18.225
35	1.225	140	19.6
40	1.6	145	21.025
45	2.025	150	22.5
50	2.5	155	24.025
55	3.025	160	25.6
60	3.6	165	27.225
65	4.225	170	28.9
70	4.9	175	30.625
75	5.625	180	32.4
80	6.4	185	34.225
85	7.225	190	36.1
90	8.1	195	38.025
95	9.025	200	40
100	10		

Your second decision is to choose what output number to report, using a screen like the one on the next page. For each unit of output reported you pay a reporting cost of 0.7. This cost is deducted from your earnings. You can choose to report any amount you like, such as your actual output or an amount less or more.

Determining who is Inspected

Once you have submitted your report the computer makes an estimate of your output. This estimate is equal to your actual output plus a random amount. The random amount has an equal chance of being any integer number between and including, -17 and 17. The random amount is equal to 0 on average, which means on average estimated output is equal to your actual output.



The computer will rank all five members in the group based on the estimated amount of output minus how much you report. The two group members with the greatest difference (gap) between the estimated amount of output and how much they report will be inspected. (Ties will be broken randomly). The other three will not be inspected that period.

Example:

Suppose you choose an output level of 20. The computer randomly selects the number +13, thus your estimated output is 33.

Your estimated difference (gap) = 33 – reported output

If you choose reported output = 20, your estimated gap is 13.

Suppose the gaps of your other four group members are: 15, 12, -1, -14.

As you are ranked second you would be inspected.

If you reported output = 22, your gap is 11 and as you are ranked third you would not be inspected.

Note: during the task when making your reporting decision you will NOT know your estimated output nor the gaps of your other group members.

What happens if you get Inspected

If you get inspected and if your actual output is greater than your reported output, you pay the following penalty, which includes the 0.7 reporting cost on the actual output that you did not report. Note that the inspection reveals your actual output.

Penalty = (0.7× Actual amount of output minus Reported output) +
(Actual amount of output minus Reported output)^2 / 40

Otherwise, you do not pay any penalty. Table 2 shows the penalty costs you would pay depending on the actual amount of output minus the reported output.

Table 2: Penalty Paid for Difference between Actual and Reported Output

Difference between Actual and Reported Output	Penalty Paid	Difference between Actual and Reported Output	Penalty Paid
0	0	21	25.73
1	0.73	22	27.50
2	1.50	23	29.33
3	2.33	24	31.20
4	3.20	25	33.13
5	4.13	26	35.10
6	5.10	27	37.13
7	6.13	28	39.20
8	7.20	29	41.33
9	8.33	30	43.50
10	9.50	31	45.73
11	10.73	32	48.00
12	12.00	33	50.33
13	13.33	34	52.70
14	14.70	35	55.13
15	16.13	36	57.60
16	17.60	37	60.13
17	19.13	38	62.70
18	20.70	39	65.33
19	22.33	40	68.00
20	24.00		

Your Earnings

As described above, each period you are given initial earnings of 20, and your overall earnings for the decision period depend upon how much you produce, how much you report, and if you are inspected, possibly a penalty.

Thus, after you have submitted your report, three things can happen:

1. You are not inspected.
2. You are inspected and your actual output is less than or equal to your reported output.
3. You are inspected and your actual output is greater than your reported output.

We summarize below how your earnings will be calculated under each scenario.

Your earnings (You are not inspected)

	20	Initial Earnings
+	Output*1	Revenue from Output
-	(Sum of output chosen by all 5 participants)^2/1000	Group Cost from Group Output
-	Amount reported*0.7	Reporting Cost
-	0	Penalty Cost
=	Period Earnings	

Your earnings (You are inspected and your actual output is less than or equal to your reported output.)

	20	Initial Earnings
+	Output*1	Revenue from Output
-	(Sum of output chosen by all 5 participants)^2/1000	Group Cost from Group Output
-	Amount reported*0.7	Reporting Cost
-	0	Penalty Cost
=	Period Earnings	

Your earnings (You are inspected and your actual output is greater than your reported output.)

	20	Initial Earnings
+	Output*1	Revenue from Output
-	(Sum of output chosen by all 5 participants)^2/1000	Group Cost from Group Output
-	Amount reported*0.7	Reporting Cost
-	0.7 × (Actual output - reported output) + (Actual output - reported output)^2 /40	Penalty Cost
=	Period Earnings	

Results

After all the members of your group have made their decisions you will see two results screens. The first screen displays your reported output, whether you were inspected and your earnings. You must write down the information displayed on your Personal Record Sheet each period.

The second screen shows the photos of all the members of your group, which were taken upon your arrival to the lab today. You will also see the reported output of all other people in your group and whether they were inspected. For the two inspected individuals you will also see their actual output, estimated output, reported output, estimated gap (which is the difference between the estimated output and the reported output) and the penalty imposed if any. Note if you are one of the two inspected people, then you would be able to see all this information about yourself.

Examples of these screens are shown on the next page.

Period

1 out 1

Remaining time [sec]: 5

Your Results

	[illegible]	Initial Earnings
+	[illegible]	Revenue from Output
-	[illegible]	Group Cost from total Group Output of [illegible]
-	[illegible]	Reporting Cost
	You were inspected	
-	[illegible]	Penalty Cost
=	[illegible]	Period Earnings

OK

Period

1 out 3

Group Summary

Photos

		You			
ID#	1	2	3	4	5
Output		[illegible]		[illegible]	
Estimated Output		[illegible]		[illegible]	
Reported Output	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Gap		[illegible]		[illegible]	
Inspected?	No	Yes	No	Yes	No
Penalty		[illegible]		[illegible]	

OK

Personal Record Sheet for Seat Number __________

Period	**1**	**2**	**3**	**…**	**…**	**23**	**24**	**25**
Initial Earnings	20	20	20	20	20	20	20	20
Revenue from Output								
Group Cost from Group Output								
Reporting Cost								
Were You Inspected? (circle one)	Yes/ No	Yes/ No	Yes/ No	Yes/ No	Yes/ No	Yes/ No	Yes/ No	Yes/ No
Penalty Cost								
Period Earnings								

Low Information Treatment: Feedback Screen

Period

1 out 1

Remaining time [sec]: 20

Group Summary

			You		
ID#	1	2	3	4	5
Output			[illegible]		
Estimated Output					
Reported Output	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Gap					
Inspected?	No	Yes	No	Yes	No
Penalty					