THE GOOD, THE BAD, AND THE REGULATOR: AN EXPERIMENTAL TEST OF TWO CONDITIONAL AUDIT SCHEMES

JEREMY CLARK, LANA FRIESEN, and ANDREW MULLER*

Conditional audit rules are designed to achieve regulatory compliance with fewer inspections than required by random auditing. A regulator places individuals into audit pools that differ in probability of audit or severity of fine and specifies transition rules between pools. Future pool assignment is conditional on current audit results. We conduct an experiment to compare two specific schemes—Harrington's Past-Compliance Targeting and Friesen's Optimal Targeting—against random auditing. We find a production possibility frontier between compliance and minimizing inspections. Optimal targeting generates the lowest inspection rates as predicted, but random auditing the highest compliance. Past-compliance targeting is intermediate. (*JEL* C91, H26, K42, L51)

I. INTRODUCTION

How can a regulatory agency achieve acceptable levels of compliance with its regulations at minimum cost of enforcement? This challenge confronts regulators in areas as diverse as tax collection, policing, customs and immigration, workplace health and safety, and natural resource management. Economists beginning with Becker (1968) have attempted to answer this question using the rational choice framework. Individuals facing a regulation will comply when the expected benefit of doing so exceeds the expected cost, and enforcement mechanisms must be set accordingly. To minimize enforcement costs, economists have proposed simple random audit regimes and, more recently, conditional audit regimes that exploit observable signals about firms or individuals. Unfortunately, it has been difficult to evaluate proposed audit schemes empirically because those who violate regulations tend to conceal their actions.

Although empirical tests have been rare, audit schemes have been tested in a steady stream of laboratory economics experiments (as surveyed in Alm and McKee [1998]). Here subjects either earn or are given increments of income, which they are then asked to disclose in order that some of the money be deducted. Motivated from the tax compliance literature, these experiments have tested the effects on compliance of alternate fines, inspection probabilities, uncertainty as to taxable income, amnesties, tax rates, and many other variables. More recently, experiments have been used to test conditional audit rules, often simple rules of thumb, that make probability of inspection dependent on subject behavior within the experiment. Though always interesting, the conditional audit rules

*We thank Peter Heffernen, John Spraggon, Paul Walker, and three anonymous referees for their helpful comments, as well as session participants at the 2001 ESA meetings in Barcelona and the 2002 World Congress of Environmental and Resource Economists in Monterey California. Financial support from the Department of Economics at the University of Canterbury is also gratefully acknowledged.

Clark: Senior Lecturer, Department of Economics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. Phone 011 643 364-2308, Fax 011 643 364-2635, E-mail jeremy.clark@canterbury.ac.nz

Friesen: Lecturer, Commerce Division, Lincoln University, P.O. Box 84, Canterbury, New Zealand. Phone 011 643 325-3627, Fax 011 643 325-3847, E-mail friesenl@kea.lincoln.ac.nz

Muller: Professor, Department of Economics, McMaster University, Hamilton, Ontario, Canada, L8S 4M4. Phone 1-905-525-9140 ext. 23831, Fax 1-905-521-8232, E-mail mullera@mcmaster.ca

ABBREVIATIONS

OT: Optimal Targeting
PCT: Post-Compliance Targeting
RAE: Random Audit Equivalent

Economic Inquiry
(ISSN 0095-2583)
Vol. 42, No. 1, January 2004, 69–87

DOI: 10.1093/ei/cbh045

tested have not generally been explicitly derived from theory and so make no claims to having been optimally designed.

This article reports on a laboratory experiment that compares inspection and compliance rates for two "forward-looking" conditional audit rules against a control rule equivalent to simple random auditing. Both rules have been designed to minimize the inspections needed to achieve target rates of compliance when regulators can vary *future* (as opposed to past) scrutiny of individuals. The two mechanisms are Past-Compliance Targeting (PCT), as proposed by Harrington (1988), and Optimal Targeting (OT), as proposed by Friesen (2003). Both schemes exploit information from current audits to assign individuals to one of two audit pools, which we dub the "green" and the "red". The green audit pool is for "good" firms and imposes a lower fine and probability of audit than the red audit pool for "bad" firms. A regulator using PCT relies on the outcomes of audits to transfer individuals in either direction between audit pools. In contrast, the OT mechanism randomly transfers individuals from the green pool to the red and uses audit outcomes only to enable compliant individuals to escape the red pool and reenter the green.

By using two audit pools, both mechanisms augment the incentives for compliance beyond the avoidance of immediate fines; noncompliance threatens greater future scrutiny, and compliance promises less. Compared to random auditing, both schemes promise to require a lower frequency of inspections to achieve a desired rate of compliance. However Harrington's PCT assumes the rules governing transfer between audit groups, whereas in Friesen's OT the transition rules are derived optimally. Thus for a given target rate of compliance, PCT should require fewer inspections in equilibrium than random auditing, but OT should require fewer still.

We find that both mechanisms do indeed succeed in lowering overall inspection rates, though only Friesen's does so significantly. However, neither mechanism achieves the overall compliance rate achieved by random auditing, and Friesen's in particular is significantly lower. If the ratio of compliance over inspection rate is taken as an ordinal measure of overall efficiency, Harrington's PCT outperforms Friesen's OT, which outperforms random auditing. Publicizing the results of audits had no significant effects in the neutral setting of the experiment.

The article is organized as follows. Section II provides a review of conditional audit rules in the tax and regulation compliance literatures. Section III describes the PCT and OT schemes in particular and the design of the experiment used to compare them. Section IV describes the results of our experiments, section V discusses the findings, and section VI provides a brief conclusion.

II. SQUEEZING BLOOD FROM A STONE—CHEAPLY

In Theory

Gary Becker (1968) wrote a seminal paper extending the rational choice model of the household to the domain of law enforcement. Individuals will violate a costly law if their analysis of the expected benefit of doing so outweighs the costs (getting caught). Allingham and Sandmo (1972) first applied this framework to tax evasion using a simple random audit rule. Here the probability of audit and fine for evasion became key parameters in the design of cost-efficient enforcement regimes. Theorists then turned their attention to using observable information supplied by taxpayers to improve cost-efficiency. In one branch, Reinganum and Wilde (1985) use a principal agent framework to propose that agencies exploit the level of income that taxpayers self-report to determine whom to audit. Agencies choose a cut-off level of reported income below which all individuals in a given class are audited.

In a second branch, theorists proposed that agencies exploit an individual's audit record when determining whom to audit. Rickard et al. (1982) proposed that the results of a person's current audit be used to determine his or her probability of back audits. Alternatively, when back audits are not possible, Landsberger and Meilijson (1982) proposed that a person's current audit outcome determine his or her probability of future audits. The latter authors showed that by targeting audits according to current audit outcomes, agencies could increase tax revenue for a given enforcement budget and fine scheme. Landsberger and Meilijson demonstrated that schemes exist that are more cost-effective than random auditing, but they did not seek to identify an optimal forward audit rule.

Greenberg (1984) extended Landsberger and Meilijson's analysis using a repeated game-theoretic approach. Greenberg proposed three audit groups, G1, G2, and G3, each with its own probability of inspection, and rules for transition that were conditional on audit status. Individuals caught underreporting income in G1 would be transferred to G2. Those caught similarly cheating in G2 would be transferred to G3, or if found in compliance, transferred back to G1. The third group serves as the ultimate deterrent, threatening certain audit and no chance of escape once entered. Greenberg found that tax evasion could be greatly reduced from that predicted under random auditing. Intuitively, tax filers should comply in G2 even with low audit rates because of the threat of transfer to G3. Unfortunately, the spectacular gains in compliance derived partly from zero discounting and unconstrained fine levels.

Harrington (1988) extended the application of Greenberg's forward-looking conditional audit rule to the realm of environmental regulation. Forward-based rules are particularly relevant for environmental regulation, where back audits of past pollution emissions or production methods may not be feasible. Harrington reduced the decision space of the regulated firms to "comply" or "violate" but also incorporated a positive discount rate and constraints on maximum fine size. These changes combined to reduce the potential efficiency gains from conditional auditing. Nonetheless, a given compliance rate could still be achieved with fewer inspections than would be needed with random auditing, even when the number of audit groups was reduced from three to two.

Harrington solved endogenously for the inspection probabilities and fines that would minimize inspections needed to induce a desired overall compliance rate. The rules governing transition between the audit groups, however, were assumed rather than solved.[1]

More recently, Friesen (2003) retained Harrington's binary decision approach for two groups but also optimized over the structure of transition rules between groups. This was shown to further reduce the inspection rate needed for a desired level of compliance. Friesen's OT scheme claims even greater cost-efficiency than Harrington's PCT scheme in equilibrium, but it holds for a narrower range of parameters. In particular, although neither scheme can be used to pursue 100% compliance, PCT can be used for higher target rates than can OT.[2] To our knowledge, the PCT and OT schemes are the only two forward-looking conditional audit rules to be formally derived.

In Practice

A limited number of empirical tests of tax compliance mechanisms have been conducted, using for example, the Taxpayer Compliance Measurement Program of the U.S. Internal Revenue Service. The limitations of these studies are discussed by Hessing et al. (1992) and Alm and McKee (1998). They must by necessity combine self-reported surveys and official records but are hampered by low sample response and attrition and confidentiality restrictions. To our knowledge, no empirical tests of the efficiency of conditional audit mechanisms have been conducted, but there is limited evidence that they are being used. In a 1999 document on innovations in its compliance policy, the U.S. Environmental Protection Agency described its objective to "work to maximize its effectiveness by strategically targeting its enforcement and compliance activities to address the most significant risks to human health and the environment" (EPA [1999], 20). The criteria for identifying firms or sectors that pose "significant risk" include compliance history, among other factors.

More formally, Helland (1998) has examined whether regulators use forward-looking audit rules in practice. He uses data from the American pulp and paper industry to test whether environmental regulators audit and fine firms according to Harrington's PCT

1. Where his transition rules required a probability of transfer, however, these were set optimally. For example, Harrington assumed that those audited in group 1 should be transferred with certainty if found in violation and kept in group 1 with certainty if found in compliance. But those who were audited in group 2 and found in compliance should have only a probability of escape back to group 1. Given the rules he assumed for transition, this probability was set optimally.

2. The cost advantage of either rule decreases as the desired compliance rate rises. When desired compliance increases beyond a critical upper bound, both schemes would need to induce compliance in both audit groups rather than just the second. When this happens, inspection costs become no cheaper than under random auditing.

model. He finds that as predicted, firms who are discovered in violation experience a one- or two-quarter period of more frequent inspections. However, he finds the basis of return to low enforcement to be self-reported violations rather than demonstrated compliance.

Moving from the limited empirical literature, experimental tests of audit mechanisms began with Friedland et al. (1978). Early studies concentrated on the effects of parameters identified by the static tax compliance model. Thus the size of fines and probability of random audit have been widely examined and found to have some effect on compliance, though less than predicted (Beck et al. [1991], Alm et al. [1992a; 1992c]). Other variables have also been considered, such as income uncertainty and risk preference (Beck et al. [1991]), the purpose of the money collected (Alm et al. [1992c]), and tax amnesties (Alm et al. [1990]). A common finding to emerge from these studies is that individuals tend to comply far more often than would be predicted in a selfish game theoretic sense, though the qualitative effects of treatments variables are usually in the direction predicted by the rational choice model. Overcompliance in experiments, as in empirical studies, has been attributed to moral or social norms (Alm et al. [1995]), or to people's tendencies to overweight small probability events such as tax audits (Alm et al. [1992c]).

Experimental tests of *conditional* audit rules began with Collins and Plumlee (1991), who test a "cut-off" audit scheme loosely based on the principal-agent model of Reinganum and Wilde (1985). A fixed number of audits were conducted on individuals reporting the lowest incomes. Collins and Plumlee also tested a "conditional cut-off" audit scheme, in which individuals are first sorted into two groups according to earning ability demonstrated in a practice session. Here the individuals reporting the lowest incomes in each group were audited. Thus the authors control the aggregate probability of inspection across regimes, though subjects could not know their individual probability of audit in either cut-off scheme. Collins and Plumlee found that both cut-off schemes were equally successful in reducing underreporting relative to random audits. Risk preferences were measured in a preexperiment questionnaire but were not found to be significant in predicting truthful reporting.

Alm et al. (1992b) provided the first test of conditional audits based on audit outcomes. They tested a forward-looking "audit reduction" scheme, though without explicit reference to prior theory. Under Alm et al.'s scheme, subjects who were audited and found in compliance would have their future probability of audit reduced from 0.04 to 0.027, and then again to 0.013. The audit probability would remain as is in the absence of audit, or revert to 0.04 if noncompliance were detected. Alm et al. found that the audit reduction scheme significantly raised compliance rates over random auditing but not as effectively as other positive inducements, such as reward lotteries for individuals found to be in compliance. When comparing compliance rates across schemes, Alm et al. imperfectly control for the ex ante probability of inspection in each.[3] Risk preferences were not controlled, though subjects' frequent all-or-nothing income reports lead the authors to conclude that risk-neutrality was a plausible assumption. In contrast to Collins and Plumlee, Alm et al.'s subjects always knew their probability of audit.

Finally, Alm et al. (1993) compared a cut-off, a backward-looking, and a forward-looking conditional audit scheme, respectively, against random auditing. Each conditional scheme was based loosely on the corresponding theories of Reinganum and Wilde (1985), Rickard et al. (1982), and Greenberg (1984). Subjects inspected in a (5%) random audit were back- (or forward-)audited for two periods with certainty if found to be underreporting income. If found to be in compliance, subjects in the forward-conditional scheme were spared the 5% chance of random audit for the next two periods. Once again, Alm et al. (1993) wrestled with the problem of comparing compliance rates across regimes with endogenously determined inspection rates. They cleverly solved this problem by running random-audit control treatments at several different levels of inspection probability. They then compared the compliance rate observed in a conditional audit treatment against the compliance rate in the control treatment with the closest matching inspection rate.

Alm et al. (1993) found that all three conditional schemes generated compliance

3. Alm et al. set the initial probability of inspection at 0.04 in every scheme, but it may fall below this level in the conditional audit reduction scheme.

significantly greater than corresponding random audit rules, though back-audits induced more compliance and fewer inspections than forward-audits. Alm et al. again simply assumed risk-neutrality and did not control for the effective discount rate as specified in the underlying theories. Subjects were in all cases informed of the audit probabilities they faced.

The experimental tests described provide encouraging evidence that conditioning audit rules on individual behaviour can increase the cost-effectiveness of regulatory enforcement. In each case, however, these experiments have been based only loosely on the mechanisms proposed in theory. We turn now to our test of two conditional audit rules, both formally derived to minimize the inspections needed to induce a target level of compliance in binary decision frameworks.

III. PCT AND OT

The Mechanisms

We consider first Harrington's (1988) PCT mechanism for two audit groups. In each decision round a firm (or taxpayer) must choose whether or not to comply with a required action that costs c. Each firm is aware of being placed in one of two audit groups, which we refer to as the green group or the more punitive red group. The probability of audit in red, p_R, is set higher than in green, p_G. Firms found in violation must pay a fine of F_G if audited while in green or of F_R if audited while in red. After each decision round transition rules determine the group a firm will be placed in for the next period. These rules follow a Markov process and are described in Figure 1. Firms found in compliance while in red are

FIGURE 1
Transition Rules in the PCT

Period t		Period t+1
If in Green and not audited	⇒	Green
If in Green and audited,		
- if found in compliance	⇒	Green
- if found in violation	⇒	Red
If in Red and not audited	⇒	Red
If in Red and audited,		
- if found in compliance	⇒	Green with Pr. ρ , Red with Pr. $1\text{-}\rho$
- if round in violation	⇒	Red

admitted back to the green group for the subsequent round with probability ρ.

Each firm's goal is minimize the present value of its expected costs over an infinite horizon. It compares the discounted present value of the following four strategies for behavior in the green and red groups, respectively: {comply, comply}, {do not comply, do not comply}, {comply, do not comply}, and {do not comply, comply}.

Harrington solves for the values of ρ, p_R, p_G, F_G and F_R that minimize the number of audits that the agency must carry out to achieve a desired overall rate of compliance. This is achieved by making it in firms' interests to pursue the strategy of never complying while in green and always complying while in red. For example, for a firm beginning in red, the expected cost of {do not comply, comply} is given by the following infinite sum:

$$(1)\qquad c + \delta[p_R\rho^* p_G F_G + (1 - p_R\rho)^* c] + \delta^2[p_R\rho([1 - p_G]p_G F_G + p_G c) + (1 - p_R\rho)(p_R\rho^* p_G F_G + [1 - p_R\rho]^* c)] + \cdots$$

The firm would pay the compliance cost c in period 0. In period 1 it would escape to green with probability $p_R\rho$, where by not complying it could expect to pay $p_G F_G$, or remain in red with probability $(1 - p_R\rho)$, where it would again pay c, and so on for subsequent periods. The general solution to (1) can be found in Friesen (2003) and can be shown to create a lower expected cost than the strategy of always complying, $c/(1 - \delta)$, never complying, $p_R F_R/(1 - \delta)$, or complying in green and not complying in red, $p_R F_R/(1 - \delta)$. Appendix Table A-1 provides the specific expected costs of each strategy in our experiment.

Friesen's (2003) OT mechanism uses a structure similar to Harrington's PCT but with the transition rules chosen along with the previous parameters to minimize audits. As Figure 1 illustrates, Harrington imposes transition probabilities between groups of either 1 or 0 for five out of six possible cases. He solves only for ρ, the probability of escape to green after being found in compliance in red. With a total of ten parameters chosen, OT results in the transition rules described in Figure 2. In OT the optimal transition probability from the green group is independent of compliance or audit status, but rather a fixed probability θ. Note that with random transfer from green to red, there is no need for audits in green ($p_G^* = 0$). Both mechanisms specify that the optimal fine for violation is $F_G = 0$ in green and $F_R = F_{MAX}$ in red. For the rest, the specific audit probabilities p_G (for the PCT), p_R, and transfer probabilities ρ and θ will depend on exogenous parameters: the target compliance rate, Z, the cost of compliance, c, maximum fine size, F_{MAX}, and discount factor δ.

Implementation

The implementation of Harrington's and Friesen's mechanisms requires a number of design decisions, which we now describe. First, we had to set the four exogenous parameters common to each model and a period income endowment. We deliberately set the target compliance rate, Z, at the moderately low level of 0.5. This is because the OT mechanism cannot be used for compliance targets near 100%. In addition, the reduction in inspections claimed by PCT relative to random auditing decreases as Z rises, making statistical discrimination difficult. Next, we set maximum fine size equal to period endowment, which in turn was set to provide subjects with average hourly earnings between 1.5–2 times the local minimum wage. We set the discount factor at $\delta = 0.9$, implemented as a probabilistic stopping rule (Davis and Holt [1993]). With a 90% probability of continuance after the first round, subjects could expect to have ten real rounds with each mechanism, though with a high variance. Finally, the cost of compliance was set to require a substantial minority of a subject's period endowment, while remaining substantially less than the fine for detected noncompliance. The exogenous and endogenous parameters used in our experiment are listed in the second and third columns of Table 1.

A second challenge for implementation of these mechanisms is that inspection and

FIGURE 2
Transition Rules in the OT

Period t		Period t+1
If in Green	⇒	Green with Pr. θ, Red with Pr. $1-\theta$
If in Red and not audited	⇒	Red
If in Red and audited,		
-if found in compliance	⇒	Green
-if found in violation	⇒	Red

TABLE 1
Parameters Used in Experiment

	Random Audit Equivalent (RAE)	Past Compliance Targeting (PCT)	Optimal Targeting (OT)
Exogenous parameters			
Endowment (points)	100	100	100
c (points)	40	40	40
F_{MAX} (points)	100	100	100
Target Z	0.5	0.5	0.5
δ	0.9	0.9	0.9
Endogenous parameters			
p_G	0	0.062296	0
p_R	0.5	0.367851	0.290173
ρ	—	0.169350	—
θ	—	—	0.709827
Equilibrium predictions			
Compliance|green	0.0	0.0	0.0
Compliance|red	1.0	1.0	1.0
Overall compliance Z	0.5	0.5	0.5
% rounds in green:	0.5	0.5	0.5
Inspection rate|green	—	p_G	—
Inspection rate|red	p_R	p_R	p_R
Overall inspection rate	0.25	0.2150734	0.1450864

compliance rates are jointly determined by subjects' decisions within the experiment. Thus it is not possible to control ex ante for, say, compliance and compare inspections rates across mechanisms. We therefore compare inspection and compliance rates simultaneously. With our target compliance rate set at 50% for both mechanisms, PCT should require an overall inspection rate of 21.5%, and OT 14.5%.

Third, it is common to compare conditional audit rules against a control treatment of simple random auditing. We have chosen a control design that is equivalent to simple random auditing, following Friesen (2003) in maximizing parallelism with the two conditional audit rules. Our control treatment retains the use of the red and green audit groups, but transition between groups becomes purely random. Audits are not carried out in the green group but are in the red group just often enough to make compliance there individually rational. As in the PCT and OT mechanisms the fine for detected violation in red is set at the maximum, $F_R = F_{\text{MAX}}$, and the cost of compliance is c. The incentives in the control treatment are such that individuals should pursue the same strategy as before—always comply in red and never comply in green. The random probability of placement in the red group is thus set equal to the overall target compliance rate, which we again set at 0.5. Although our random audit control should thus achieve the same overall compliance rate as PCT and OT, it should require a higher inspection rate of 25%.

With these design decisions made implementation is relatively straightforward. We employ a within-subject design to maximize the power of statistical tests, as our predicted inspection rates are not that far apart. Order effects are addressed by running sessions in all possible sequences: ABC, ACB, BAC, BCA, CAB, and CBA. Neutral language is used throughout. In a given decision round of a mechanism, subjects choose between Option A (compliance) and Option B (violation), and then face a possible audit described as "entering a random draw." Regarding risk preference, both the PCT and OT mechanisms assume risk-neutrality. Rather than presume this risk preference, we attempt to induce risk neutrality by having the compliance decision made over lottery tickets (as in Davis and Holt [1993]). Each subject begins a decision round endowed with 100 points, and the cost of compliance is the surrender of 40 such points to the experimenter. These points are used to enter a random draw for \$1 each round. By surrendering 40 points, a subject reduces his or her probability of winning the \$1 draw from 100% to 60%, or by \$0.40 on average.

The combination of a probabilistic stopping rule, within-subject design, and payout over lottery tickets could result in a very complex environment for subjects to understand. This in turn could result in less meaningful compliance decisions. We thus take several steps to aid comprehension. First, we distribute a paper color-coded schematic diagram to each subject for each mechanism as we progress through the experiment. These are reproduced in Appendix Figures B-1 through B-3, absent the color. Second, we give subjects ten hypothetical practice rounds with each mechanism prior to its first real round. So the subjects in a particular session might experience, for example, 10 practice and 7 real rounds of the PCT, then 10 practice and 14 real rounds of the control, and finally 10 practice and 9 real rounds of the OT.

Finally, we altered a parameter within our overall design so as to make all three mechanisms' optimal strategies more transparent. In particular, we calculated endogenous parameters using a cost of compliance, c, of 50 points for all three mechanisms, but then reduced this cost to 40 points. Why? All three mechanisms extract maximum efficiency by setting parameters so that the strategy {do not comply in green, comply in red} *just* dominates the strategy of never complying. By slightly reducing the cost of compliance, we increase the payoff dominance of the optimal strategy. We hope in this way to reduce decision errors but at the expense of lowering slightly the potential efficiency of all three mechanisms. Put another way, if we had believed that indifferent subjects would always comply when in red, we could have set inspection, audit, and transfer probabilities differently for each regime so as to predict even lower inspection rates for all three mechanisms.

Information Effects

A second treatment variable addressed in our experiment is the effect of publicizing others' audit results on compliance rates and cost efficiency. Real-world publicity threatens firms or taxpayers with shame from being exposed in audits but also provides better information on the compliance strategies being pursued by others. In the neutral-language setting we adopt, only the information effect can be captured. We do this by running all sessions (in all orders) in high- and low-information versions. In the high-information treatment, subjects are informed after each decision round of the number of people who have been in the green and red groups, how many of these have been audited, and how many of the audited have complied or not complied. In the low-information treatment this information is withheld. With 2 information levels and 6 possible mechanism orders, we run 12 sessions in total.

IV. RESULTS

We ran 12 complete and 1 partial session of the experiment over a three-month period in March to May 2001. Overall, 141 subjects took part in the complete sessions (where each was exposed to all three mechanisms), and 12 took part in a partial session that unexpectedly crashed after completing PCT. Subjects were recruited from large first- and second-year classes in economics, mathematics, and political science at the University of Canterbury in Christchurch, New Zealand. Each session lasted between 80 and 120 minutes, and subjects earned NZ$22.13 on average. (The New Zealand minimum wage was updated in 2000 to $7.55/hour.)

Tables 2 and 3 provide a summary of our results for compliance and inspection rates, respectively. In all cases we take as our unit of observation each person's compliance or inspection outcomes averaged over all decision rounds under a given regime. The Mann-Whitney nonparametric test of differences in frequency distribution indicated no significant difference between high- and low-information treatments for any regime. Hence the results are pooled in the table and for all subsequent analysis. Note that the sample sizes differ because of the additional partial session for PCT and because some subjects did not experience both audit groups in a given regime.[4]

4. Half of the subjects were placed in the red and green groups for the first real round of each mechanism, and they were informed of this. Placement was varied across mechanisms so that no subject was consistently placed in one group, and every possible sequence was equally represented.

Order Effects

Tests for order effects in compliance and inspections were carried out using the Kruskal-Wallis test for differences between independent samples. No order effects were found for inspection rates under any audit regime. Order effects did emerge, however, for some compliance rates under PCT and random audit equivalent (RAE). For the PCT, Table 2 shows that subjects were less likely to comply in green and more likely to comply in red the later they experienced the regime in the sequence of three. This effect was significant in both audit pools (p-value$_{\text{Green}} = 0.038$, p-value$_{\text{Red}} = 0.020$). Interestingly, these contrasting compliance trends in red and green offset each other so that no significant order effect was found for overall compliance (p-value$_{\text{Overall}} = 0.357$). For the RAE, order effects were found for compliance in green (p-value$_{\text{Green}} = 0.052$) and more ambiguously in red (p-value$_{\text{Red}} = 0.070$). In contrast, no order effects in compliance were observed for OT.

The order effects in compliance in PCT are suggestive of learning that comes specifically from being able to compare features across audit regimes.[5] They create potential difficulties for pooling compliance observations under this regime. Our approach will be to persevere in pooling and then consider the implications order effects have for the results.

Treatment Effects

We can evaluate each audit regime's performance (1) absolutely, against its own theoretical prediction, or (2) relatively, against the performance of the other regimes. Absolute performance in compliance and inspection rates is reported in the final column of Tables 2 and 3, respectively, and may be compared with the predictions made in Table 1. Most results line up reasonably well with predictions. Formal comparisons are made with t tests, and significant differences are indicated on Tables 2 and 3 with asterisks.[6] Recall that subjects

5. Tests for order effects and learning *within* a regime are difficult to make, as subjects in different sessions experienced a given regime or audit pool for very different numbers of real rounds. (All experienced ten practice rounds with each regime.) This is why our unit of observation is the average behavior of a subject under a given regime and audit pool.

6. Although the distribution generating the underlying compliance or inspection observations for these variables

TABLE 2
Compliance Rates Observed under Each Regime

		Order of Presentation			
		First	**Second**	**Third**	**Pooled**
A. When in green group					
Random audit equivalent:	Mean	0.167	0.072	0.029	0.089**
	N	41	43	41	125
Past compliance targeting:	Mean	0.175	0.108	0.062	0.117**
	N	40	32	36	108
Optimal targeting:	Mean	0.074	0.037	0.097	0.067**
	N	41	39	31	111
3 combined:	Mean	0.138	0.070	0.059	
	N	122	114	108	
B. When in red group					
Random audit equivalent:	Mean	0.751	0.843	0.754	0.782**
	N	41	39	41	121
Past compliance targeting:	Mean	0.691	0.749	0.845	0.754**
	N	41	27	30	98
Optimal targeting:	Mean	0.588	0.635	0.608	0.610**
	N	45	41	34	120
3 combined:	Mean	0.674	0.740	0.733	
	N	127	107	105	
C. Overall compliance rates					
Random audit equivalent:	Mean	0.465	0.442	0.430	0.446
	N	47	47	47	141
Past compliance targeting:	Mean	0.428	0.363	0.374	0.391**
	N	59	47	47	153
Optimal targeting:	Mean	0.336	0.356	0.303	0.332**
	N	47	47	47	141
3 combined:	Mean	0.411	0.387	0.369	
	N	153	141	141	

*,**For pooled results, indicates significant difference from theoretical prediction at the 5% and 1% levels, respectively. *t* tests calculated on SPSS version 10.0.

should never comply when in green: the average of individual average compliance rates ranged between 6.7% and 11.6% across the three regimes. A greater divergence from theory was observed in compliance rates in the red group. Though all subjects should comply when in red, actual compliance ranged from only 61.0% under Friesen's OT to 75.4% under Harrington's PCT and 78.2% under the RAE.

Turning to inspection rates, those generated for subjects in each audit group by the computer random number generator were generally as expected, though there seemed to be unusually many inspections ex post in the green group in Harrington's PCT and in the red group in Friesen's OT. Overall inspection rates depend on subjects' compliance decisions, and through them, time spent in each audit pool.[7] The overall rates were not significantly different than predicted for RAE or Harrington's PCT but were slightly higher than predicted for Friesen's OT (17.8% rather than 14.5%), as undercomplying subjects spent "too much" time in red.

We turn next to relative comparisons across regimes. Does Harrington's PCT achieve the same level of compliance as RAE, with fewer

may not be normal, the sample mean for each individual's inspection or compliance rate should be distributed normally in large samples. Hence the *t* test is appropriate.

7. Subjects were predicted to spend 50% of rounds in the red group for all three regimes. In fact they spent 49.7% of rounds in red under RAE, 46.4% in red under PCT, but 56.8% in red under OT. Only the OT difference is significant at the 5% level and was caused by undercompliance.

TABLE 3
Inspection Rates Observed under Each Regime

		Order of Presentation			
		First	**Second**	**Third**	**Pooled**
A. When in green group					
Random audit equivalent:	Mean	—	—	—	—
	N				
Past compliance targeting:	Mean	0.108	0.074	0.135	0.105*
	N	40	32	36	108
Optimal targeting:	Mean	—	—	—	—
	N				
3 combined:	Mean	—	—	—	
	N				
B. When in red group					
Random audit equivalent:	Mean	0.508	0.493	0.478	0.493
	N	41	39	41	121
Past compliance targeting:	Mean	0.368	0.445	0.400	0.399
	N	41	27	30	98
Optimal targeting:	Mean	0.351	0.304	0.412	0.352*
	N	45	41	34	120
3 combined:	Mean	0.407	0.408	0.434	
	N	127	107	105	
C. Overall inspection rates					
Random audit equivalent:	Mean	0.261	0.243	0.227	0.244
	N	47	47	47	141
Past compliance targeting:	Mean	0.207	0.194	0.188	0.197
	N	59	47	47	153
Optimal targeting:	Mean	0.190	0.161	0.184	0.178*
	N	47	47	47	141
3 combined:	Mean	0.218	0.199	0.200	
	N	153	141	141	

*,**For pooled results, indicate significant difference from theoretical prediction at the 5% and 1% levels, respectively. *t* tests calculated on SPSS version 10.0.

inspections? Does Friesen's OT require still fewer inspections? Descriptive comparisons can be made by moving up and down the final column of Tables 2 or 3. Formal comparisons are made using the Wilcoxon signed rank test of paired samples and presented in Table 4. The partial 13th session is necessarily omitted for these within-subject tests. Note that sample size again varies for paired observations because not all subjects experienced a given audit pool in every regime.

The comparative results of Table 4 provide arguable evidence that Harrington's PCT mechanism outperforms the control and Friesen's OT, though on different dimensions. Regarding compliance, the PCT does not achieve significantly less than RAE in (1), and achieves significantly more than OT in (3), both when subjects are in the red group and overall. Regarding inspection rates, the PCT generated significantly fewer inspections for subjects in the red group than the RAE (see [1]). However, the PCT requires inspections for subjects in the green group and the RAE does not, so that the PCT's overall inspection rate is suggestively but not significantly lower (p-value $= 0.168$, two-tailed.). At the same time, the PCT does not require significantly more inspections than Friesen's OT (in [3]). A run-off comparison between OT and RAE is more problematic. As promised in theory, the OT requires significantly fewer inspections than RAE in the red group as well as over all (neither regime requires inspections in the

TABLE 4
A Signed-Rank Test Comparison of Regimes

	(1) PCT-RAE	(2) OT-RAE	(3) OT-PCT
		Standard normal Z values	
Compliance rate | green:	1.056	−0.665	−0.309
	($N=87$)	($N=97$)	($N=73$)
Compliance rate | red:	−1.053	−3.688***	−1.909*
	($N=76$)	($N=102$)	($N=74$)
Overall compliance:	−1.107	−3.024***	−1.942**
	($N=141$)	($N=141$)	($N=141$)
% rounds in green:	0.730	−1.865*	−2.028**
	($N=141$)	($N=141$)	($N=141$)
Inspection rate | green:[a]	—	—	—
Inspection rate | red:	−1.915*	−3.720***	−1.217
	($N=76$)	($N=102$)	($N=74$)
Overall inspection:	−1.378	−2.505***	−0.589
	($N=141$)	($N=141$)	($N=141$)

[a]No inspections were carried out in RAE or OT.
*, **, ***denote significance at the 10%, 5% and 1% levels, respectively. Calculated on SPSS version 10.0.

green group). However the OT achieved a surprisingly lower rate of compliance than the RAE in the red group and over all.

A single cardinal measure of the compliance-inspection trade-off between regimes is not possible because while the expected compliance cost per round is $0.40, the cost to the agency per inspection is unspecified. However an ordinal measure of regime performance, average compliance over average inspection, (C/I), can be calculated for each regime at the session level and then overall.[8] For a given session *j*,

$$C/I_{\text{Session } j} = (1/n_{Session\ j}\Sigma C_i)/(1/n_{Session\ j}\Sigma I_i) = \Sigma C_i/\Sigma I_i, \quad (2)$$

where C_i and I_i refer to the average compliance and inspection rates, respectively, of individuals within session *j*. The overall *C/I* is the weighted mean of the C/I_{Session} ratios, where the weights reflect the number of participants per session (11 or 12). This yields a *C/I* index for Harrington's PCT of 2.21, for Friesen's OT of 1.94, and for the control RAE of 1.90. If the cost per inspection were equal to the cost per compliance, this would indicate that OT's inspection rate advantage over the control RAE more than compensates for its lower compliance. But Harrington's PCT would dominate both.

The preceding comparisons are muddied somewhat by order effects found in the PCT for red and green groups and in the RAE for the green group. These order effects are suggestive of learning, as subjects in these cases behaved more "rationally" as they accumulated experience across regimes. Fortunately, the lack of order effects on compliance rates in OT provides a convenient benchmark. Red group average compliance under pooled OT was lower than that under PCT in whichever order PCT appeared. Thus the finding that red group compliance was lower in OT than in PCT appears robust to order effects. By extrapolation, the more experience subjects gained across regimes, the greater this disparity would grow.

Untangling order effects for green group compliance results is less straightforward. Average green compliance in both PCT and RAE among "inexperienced" subjects was markedly higher than in pooled OT. This

8. *C/I* ratios may also be constructed at the individual level but are undefined for individuals who are never inspected.

higher compliance disappeared and reversed, however, as subjects gained experience across regimes. Combining results, PCT red compliance rose above that in OT, and PCT green compliance fell below it. As a result, the overall compliance in PCT remained steadily superior to that in OT. Meanwhile, RAE compliance advantage in green over OT disappeared with experience but not dramatically enough to change RAE's overall compliance advantage over OT.

V. DISCUSSION

Our experimental test of Harrington's PCT and Friesen's OT against random auditing yielded several results consistent with theoretical predictions. Both conditional audit mechanisms required fewer inspections than equivalent random auditing, though only OT generated a significant reduction. OT also required fewer inspections than PCT, but the difference was not statistically significant. Theory was also correct in predicting that subjects in all three mechanisms would comply more often in the punitive red group than when they were in green.

Other results, however, were less anticipated. Informing subjects of the audit outcomes of others did not seem to affect compliance rates, all else equal. This held across all three audit mechanisms. In the neutral setting of our experiment, this suggests that the information effect of audit publicity does not assist subjects in making their own compliance decisions. It would be interesting to see if audit publicity would augment compliance incentives in a nonneutral setting where moral value or social approval could be attached to individual compliance decisions.

More surprisingly, all three mechanisms failed to induce full compliance among subjects when they were placed in the punitive red group. This failure was especially pronounced under OT, where subjects complied only 61% of the time on average when in red, as compared with 75% of the time in PCT and 78% of the time under random auditing. There was also a converse overcompliance in green on average across all three mechanisms, but to a much lesser extent. As a result, the *overall* rate of compliance under the two conditional audit mechanisms was significantly less than predicted and was only just statistically indistinguishable under the RAE control.

This general undercompliance of subjects when in red in both PCT and OT had flow-on effects. Although inspection rates *within* a group were determined by fixed rules or by random draw, *overall* inspection rates were affected by where subjects were spending most of their time. Thus, subjects who undercomplied in red were detained there more often than would be predicted by theory.[9] This in turn necessitated more inspections, which eroded the savings in inspection rates promised under both conditional mechanisms. In particular, OT succeeded where the PCT failed in achieving significantly lower inspection rates than random auditing. But its inspection rate was still significantly higher than predicted.

These results raise the question: Why did subjects undercomply when in the red audit group, particularly under OT? We take several approaches to answer this question.

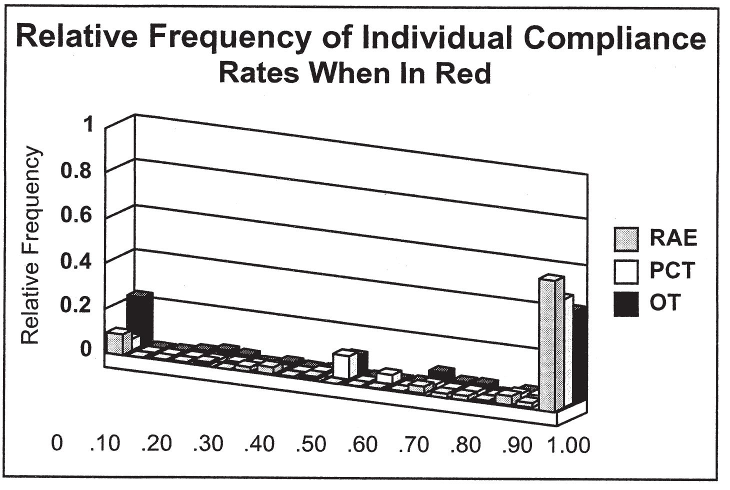
First, some insight can be gained from examining behavior at the individual level. Figure 3 shows the distribution of individual compliance rates in red across the three mechanisms. A full 60% of subjects chose to comply 100% of the time when in the control, whereas both PCT and OT in particular induced a greater dispersion of compliance rates. In particular, almost 22% of subjects *never* complied in the red group under the OT regime.

If we examine subjects' decisions in one audit group conditional on their behavior in the other, we can look for evidence that general undercompliance is caused by individual-specific characteristics. For example, if some subjects consistently comply in both groups and others consistently do not, then general undercompliance might derive from the distribution of compliant and noncompliant types in our experiment.[10] Conversely, if some subjects always comply in red and not in green, and others take an intermediate path in both groups, then overall undercompliance might derive from the distribution of rational and confused participants.

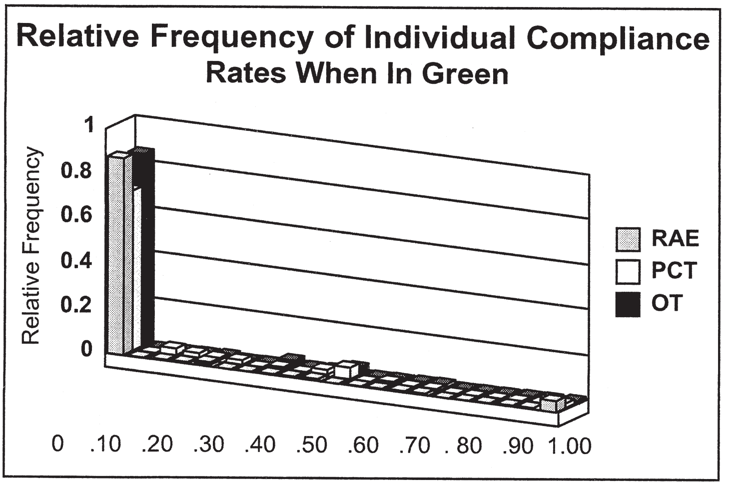
9. The unusually high number of inspections generated randomly in red under OT would not in itself detain subjects in red. Given that subjects were not complying, a low or average inspection rate would also have detained them there.

10. Because neutral language is used in the experiment, these would more properly be described as Option A and Option B types.

FIGURE 3
The Distribution of Average Compliance Rates



(A)



(B)

For each mechanism, we calculated four such conditional measures, described in Appendix Table C-1. To take an example, consider the 90 individuals who experienced both the green and red groups under OT. We conditioned this sample on having complied less or more than 60% of the time while in red and then examined the corresponding compliance of these subjects when in green. The 28 least compliant subjects when in red had an average compliance rate in green of 6.8%. This turns out to be virtually indistinguishable from the choices of the 62 subjects who were more compliant in red; their average compliance in green was 5.6%. Thus, those who could be classified as noncompliers in red were neither more nor less prone to comply in green. More generally, we found no significant evidence that an individual's behavior in one group could be used to predict his or her behavior in the other in any of the mechanisms.

A second explanation for the various degrees of undercompliance in red is that subjects were generally confused as to the dynamic aspects of each audit rule, and so myopically concentrated on the immediate probabilities of audit and fine. Consistent with this hypothesis, red group compliance was highest under RAE, where the immediate expected fine of $0.50 clearly exceeded the cost of compliance ($0.40). Red compliance was intermediate under PCT, where the immediate expected fine fell below compliance cost, to $0.37, and it was lowest under OT, where the immediate expected fine fell further to $0.29. A problem with this explanation, however, is that we would expect that myopia driven by confusion would be stronger and more enduring under more complex mechanisms. Our test of order effects, however, found that subjects had the greatest tendency to converge toward the optimal strategy under the PCT and the least convergence under OT. It is our conjecture that the OT is, if anything, *less* complex than the PCT, because the consequence of compliance in red involves a simple (rather than compound) lottery.

A third explanation for undercompliance under OT relative to PCT comes from a theoretical comparison of the two mechanism's structures. OT and PCT have three design differences that are supposed to combine to keep compliance rates identical. The PCT threatens inhabitants of the red group with a higher probability of immediate audit (36.8%) and makes the green group a more tempting escape destination because of its low audit rate. On the other hand, the PCT only offers a low 6.2% chance of escape from red (36.8% × 16.9%) as a reward for compliance. Friesen's OT has a lower threat of immediate audit in red (29.0%), and makes the green group a less tempting escape destination because of its high rate of random transfer back to red. But OT also offers a much higher chance of escape from red as a reward for compliance (29.0%). We could speculate that the PCT's offer of escape from a more punitive prison to a more tempting destination is more effective in inducing compliance, even though that offer comes with a lower probability. This would be true particularly if subjects tended to overestimate their probability of escape to green under PCT because of its compound lottery.

A final explanation for general undercompliance in red relates to weak payoff dominance. As discussed under implementation, both the PCT and OT extract maximum savings in inspection rates by making the strategy {do not comply, comply} in green and red *just* dominate the strategy {do not comply, do not comply}. Recognizing this possibility in advance, we implemented lower compliance costs than those we used to calculate inspection and transition probabilities. This increased the payoff dominance of the optimal strategy to the levels demonstrated in Appendix Table A-1. Even so, subjects could remain prone to mistakenly adopting suboptimal strategies that involve noncompliance in red. Mistaken compliance in green would be less frequently observed, because in most cases the strategies that include it are more strongly dominated. Note that payoff dominance can explain the undercompliance in red observed across all three mechanisms but not why it was worse under OT than PCT. For as Appendix Table A-1 makes clear, the loss from the second-best strategy of uniform noncompliance was actually greater under OT than under PCT.

Policy makers wishing to field test or implement the OT or PCT mechanisms should take note of their weak payoff dominance properties. As designed, both mechanisms make individuals almost indifferent between complying when in red and never complying. Our results suggest that even with moderate decision error costs, individuals will undercomply in red, spend more time there, and thus require more inspections than predicted. Regulators could increase payoff dominance as we did by implementing the PCT or OT mechanism as if compliance costs were higher than they are thought to be, or by raising inspection rates in red above what the theory requires. Unfortunately, either step would also lower the savings in inspection rates that either mechanism could offer over random auditing.

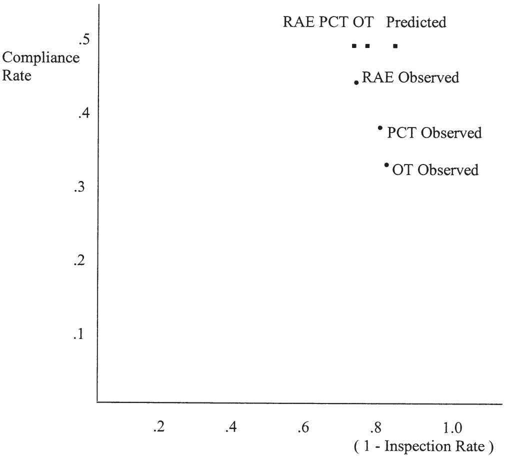
VI. CONCLUSION

Harrington's PCT and Friesen's OT are forward-looking conditional audit rules designed to minimize the inspections regulators must make to achieve a target rate of compliance. These rules exploit an observable characteristic of tax payers or firms—their current audit record—to assign individuals to differing audit groups for future periods. Transition rules between audit groups can augment the stick for present compliance (avoiding fines) with the carrot of future placement in preferable audit groups. By placing fewer restrictions on the optimal design of transition rules, the OT claims to require even fewer inspections than the PCT.

Conditional audit rules have attracted some criticism. Harford and Harrington (1991) observe that the objective of minimizing inspection costs conflicts with minimizing the private cost of compliance to firms because marginal compliance costs will differ in equilibrium for otherwise identical firms. Harford (1991) shows, however, that the net *social* gains are likely to be positive in many cases, particularly where the marginal cost of compliance is close to constant.

In theory, both mechanisms should achieve a given level of compliance with fewer inspections than random auditing, with Friesen's OT requiring even fewer than Harrington's PCT. Our results suggest rather that enforcement agencies may instead face a production possibility frontier between compliance and minimized inspection, as illustrated in Figure 4. Random auditing seems most effective at achieving compliance but at a high cost in inspection rates. OT seems most effective in minimizing inspection rates but at a cost in the compliance obtained. PCT achieves (almost) as much compliance as random auditing while requiring almost as few inspections as OT.

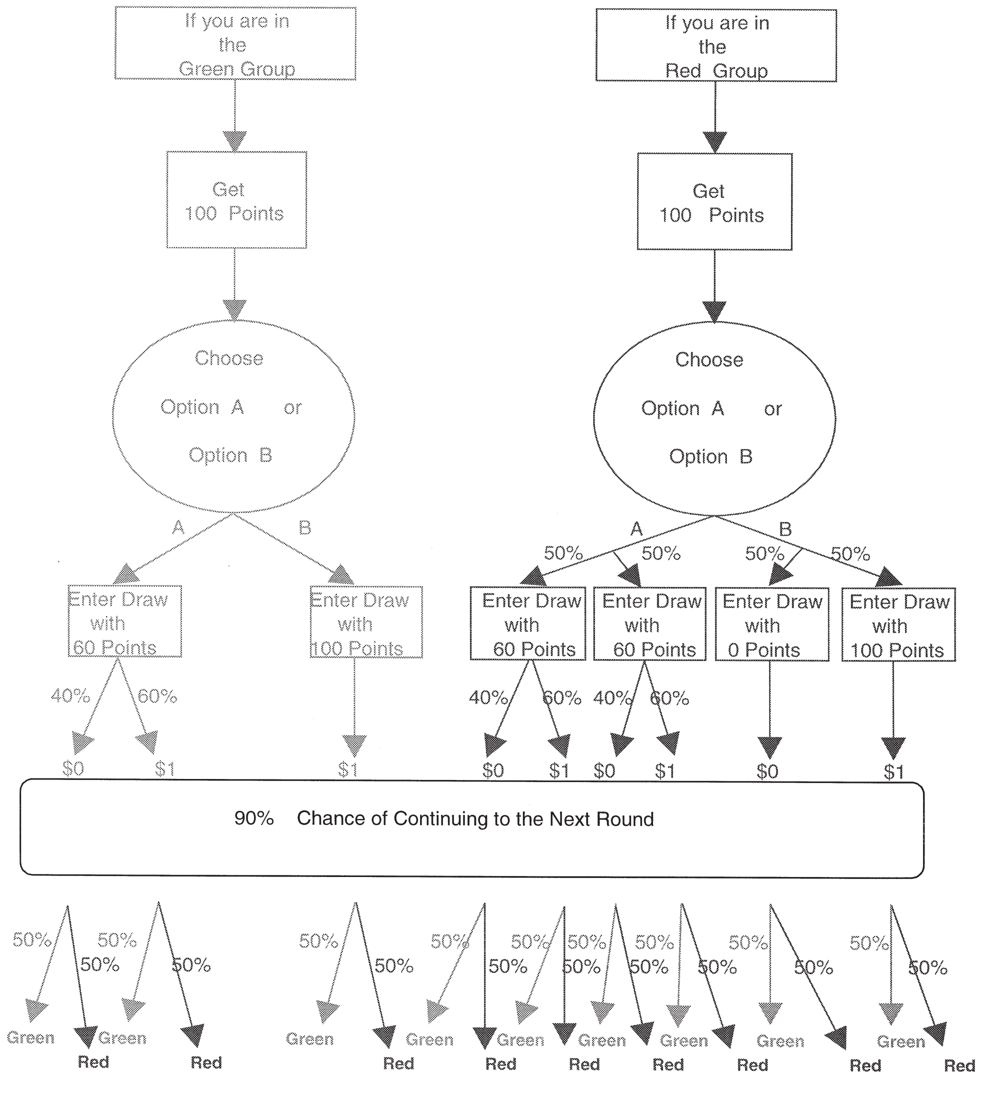
FIGURE 4
The Inspection–Compliance Trade-off Observed



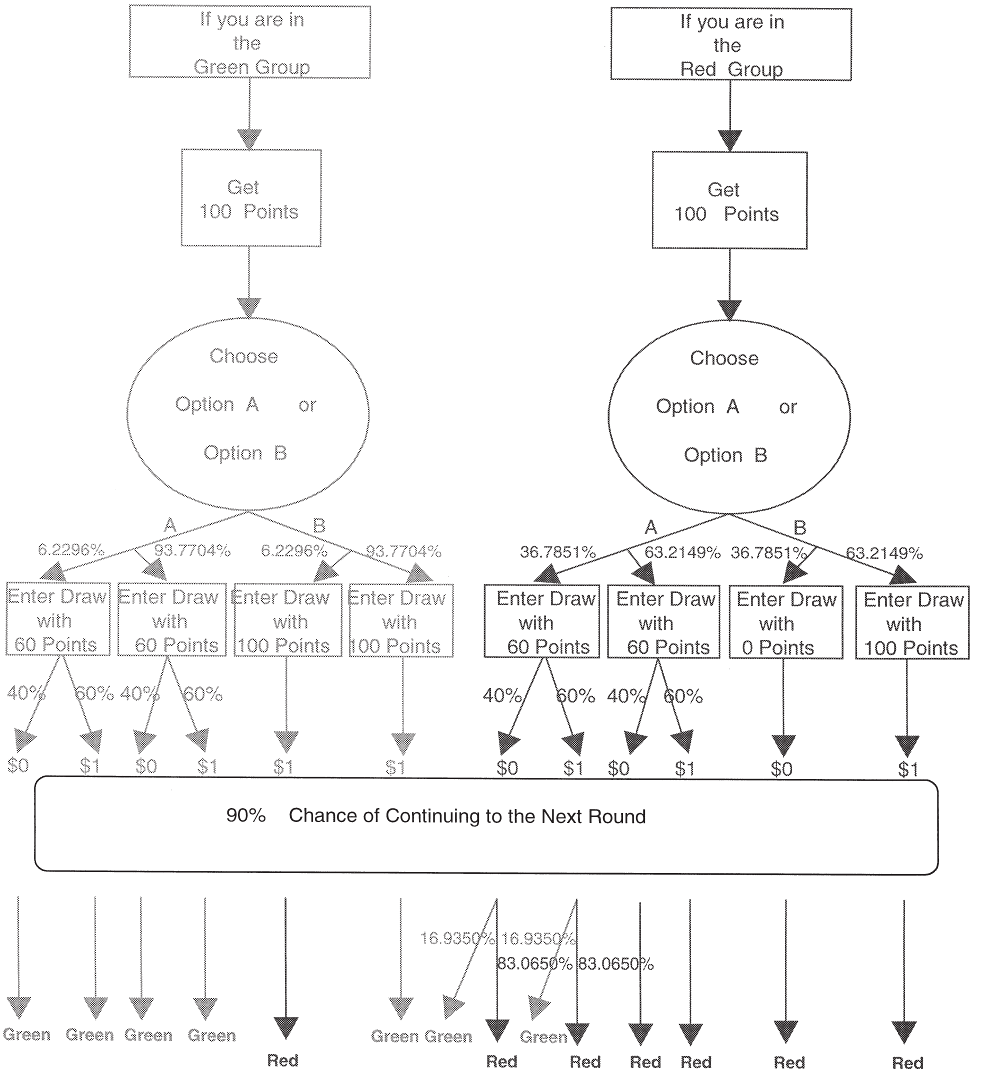
APPENDIX TABLE A-1
Expected Payoff from Alternative Compliance Strategies in Our Experiment

Mechanism	Strategy: Green Group	Strategy: Red Group	Expected Payoff If Begin in Green	Expected Payoff If Begin in Red
RAE	Comply	Comply	\$10 – \$4.00 = \$6.00	\$10 – \$4.00 = \$6.00
	Do not comply	**Comply**	**\$10 – \$1.80 = \$8.20**	**\$10 – \$2.20 = \$7.80**
	Comply	Do not comply	\$10 – \$4.45 = \$5.55	\$10 – \$4.55 = \$5.45
	Do not comply	Do not comply	\$10 – \$2.25 = \$7.75	\$10 – \$2.75 = \$7.25
PCT	Comply	Comply	\$10 – \$4.00 = \$6.00	\$10 – \$4.00 = \$6.00
	Do not comply	**Comply**	**\$10 – \$1.03 = \$8.97**	**\$10 – \$2.92 = \$7.08**
	Comply	Do not comply	\$10 – \$4.00 = \$6.00	\$10 – \$3.70 = \$6.30
	Do not comply	Do not comply	\$10 – \$1.30 = \$8.70	\$10 – \$3.70 = \$6.30
OT	Comply	Comply	\$10 – \$4.00 = \$6.00	\$10 – \$4.00 = \$6.00
	Do not comply	**Comply**	**\$10 – \$1.68 = \$8.32**	**\$10 – \$2.32 = \$7.68**
	Comply	Do not comply	\$10 – \$3.20 = \$6.80	\$10 – \$2.90 = \$7.10
	Do not comply	Do not comply	\$10 – \$2.10 = \$7.90	\$10 – \$2.90 = \$7.10

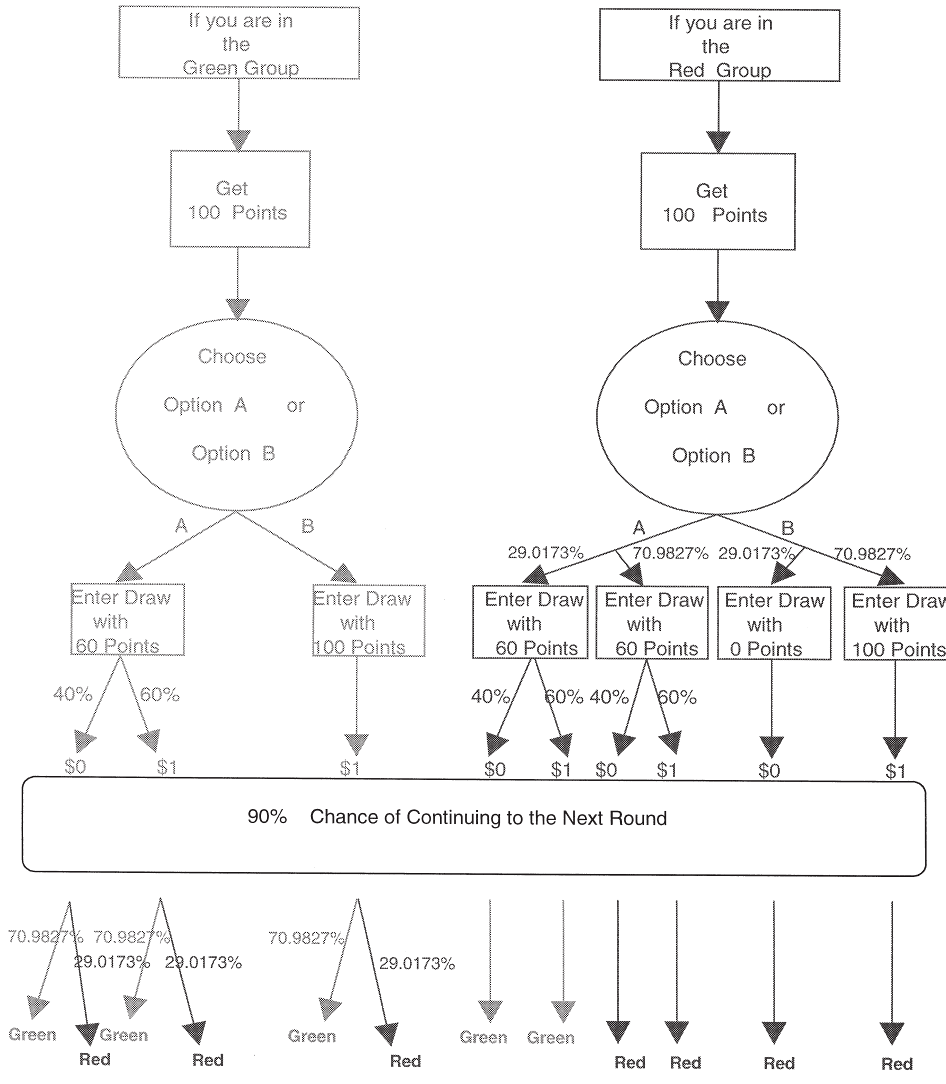
APPENDIX FIGURE B-1
Diagram of Random Auditing Equivalent (RAE)



APPENDIX FIGURE B-2
Diagram of Past-Compliance Targeting (PCT)



APPENDIX FIGURE B-3
Diagram of Optimal Targeting (OT)



APPENDIX TABLE C-1
Conditional Mean Compliance Rates

	RAE			PCT			OT		
	Compliance In Green	**Compliance In Red**	***N***	**Compliance In Green**	**Compliance In Red**	***N***	**Compliance In Green**	**Compliance In Red**	***N***
Very compliant vs. others									
If compliance in red >0.90	0.084	—	60	0.107	—	37	0.042	—	44
If compliance in red $\leq$ 0.90	0.092	—	45	0.089	—	16	0.077	—	46
If compliance in green >0.35	—	0.759	10	—	0.890	6	—	0.621	7
If compliance in green $\leq$ 0.35	—	0.758	95	—	0.874	47	—	0.725	83
Very noncompliant vs. others									
If compliance in red <0.60	0.086	—	27	0.139	—	6	0.068	—	28
If compliance in red $\geq$ 0.60	0.088	—	78	0.097	—	47	0.056	—	62
If compliance in green <0.10	—	0.760	92	—	0.878	41	—	0.720	80
If compliance in green $\geq$ 0.10	—	0.743	13	—	0.868	12	—	0.688	10

Note: None of the conditional mean compliance rates was found to be significantly different from its compliment, using two sided *t* tests with equality of variance not assumed.

REFERENCES

Alm, J., and M. McKee. "Extending the Lessons of Laboratory Experiments on Tax Compliance to Managerial and Decision Economics." *Managerial and Decision Economics*, 19, 1998, 259–75.

Alm, J., Jackson, B., and M. McKee. "Estimating the Determinants of Taxpayer Compliance with Experimental Data." *National Tax Journal*, 45(1), 1992a, 107–14.

———. "Deterrence and Beyond: Toward a Kinder, Gentler IRS," in *Why People Pay Taxes*, edited by J. Slemrod. Ann Arbor: University of Michigan Press, 1992b.

Alm, J., G. McClelland, and W. Schulze. "Why Do People Pay Taxes?" *Journal of Public Economics*, 48, 1992c, 21–38.

Alm, J., Cronshaw, M., and M. McKee. "Tax Compliance with Endogenous Audit Selection Rules." *Kyklos*, 46(1), 1993, 27–45.

Alm, J., McKee, M., and W. Beck. "Amazing Grace: Tax Amnesties and Compliance." *National Tax Journal*, 43, 1990, 23–27.

Alm, J., Sanchez, I., and A. de Juan. "Economic and Noneconomic Factors in Tax Compliance." *Kyklos*, 48(1), 1995, 3–18.

Allingham, M., and A. Sandmo. "Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics*, 1, 1972, 323–38.

Beck, P., Davis, J., and W. Jung. "Experimental Evidence on Taxpayer Reporting under Uncertainty." *Accounting Review*, 66(3), 1991, 535–58.

Becker, G. "Crime and Punishment: An Economic Approach." *Journal of Political Economy*, 76, 1968, 169–217.

Collins, J., and R. D. Plumlee. "The Taxpayer's Labor and Reporting Decision: The Effect of Audit Schemes." *Accounting Review*, 66(3), 1991, 559–76.

Davis, D., and C. Holt. *Experimental Economics*. Princeton: Princeton University Press, 1993.

Environmental Protection Agency. "Protecting Your Health and the Environment through Innovative Approaches to Compliance: Highlights from the Past 5 Years." EPA/300-K-99-001, Washington, DC, 1999.

Friedland, N., S. Maital, and A. Rutenberg. "A Simulation Study of Income Tax Evasion." *Journal of Public Economics*, 10, 1978, 107–16.

Friesen, L. "Targeting Enforcement to Improve Compliance with Environmental Regulations." *Journal of Environmental Economics and Management*, 46, 2003, 72–85.

Greenberg, J. "Avoiding Tax Avoidance: A (Repeated) Game-Theoretic Approach." *Journal of Economic Theory*, 32, 1984, 1–13.

Harford, J. "Measurement Error and State-Dependent Pollution Control Enforcement." *Journal of Environmental Economics and Management*, 21, 1991, 67–81.

Harford, J., and W. Harrington. "A Reconsideration of Enforcement Leverage When Penalties Are Restricted." *Journal of Public Economics*, 45, 1991, 391–95.

Harrington, W. "Enforcement Leverage When Penalties Are Restricted." *Journal of Public Economics*, 37, 1988, 29–53.

Helland, E. "The Enforcement of Pollution Control Laws: Inspections, Violations, and Self-Reporting." *Review of Economics and Statistics*, 80(1), 1998, 141–53.

Hessing, D., H. Elferrs, H. Robben, and P. Webley. "Does Deterrence Deter? Measuring the Effect of Deterrence on Tax Compliance in Field Studies and Experimental Studies," in *Why People Pay Taxes*, edited by J. Slemrod. Ann Arbor: University of Michigan Press, 1992.

Landsberger, M., and I. Meilijson. "Incentive Generating State Dependent Penalty System." *Journal of Public Economics*, 19, 1982, 333–52.

Reinganum, J., and L.Wilde. "Income Tax Compliance in a Principal-Agent Framework." *Journal of Public Economics*, 26, 1985, 1–18.

Rickard, J., Russell, A., and Howroyd, T. "A Tax Evasion Model with Allowance for Retroactive Penalties." *Economic Record*, 58, 1982, 379–85.