

Tax compliance as a coordination game[☆]

James Alm^{a,*}, Michael McKee^{a,b}

^a *Department of Economics, Andrew Young School of Policy Studies, Georgia State University,
Atlanta, GA 30303-3083, USA*

^b *Department of Economics, University of Tennessee, Knoxville, TN, USA*

Received 21 December 2001; accepted 12 February 2003

Available online 31 December 2003

Abstract

This paper uses laboratory experiments to investigate compliance behavior when returns are selected for audit based upon the deviation of each individual's tax report from the average report of all other taxpayers. Our experimental results indicate that individuals find it difficult to coordinate on the zero-compliance equilibrium. However, pre-game communication that mimics information-sharing provided by tax guides provides a mechanism that allows such coordination. Nevertheless, the tax authority is able to overcome this taxpayer coordination by a subtle change in its audit rule, a change that targets the audits in a different way without increasing the number of audits.

© 2003 Elsevier B.V. All rights reserved.

JEL classification: H2

Keywords: Tax compliance; Coordination game

1. Introduction

A major tool of tax agencies to reduce income tax evasion is the audit of individual tax returns. With an audit, the amount of income (or other tax return items) reported by the taxpayer is checked by the agency, and fines are imposed on the taxpayer, if he or she is found to have under-paid the tax. The simplest method of selecting returns is a random

[☆] The experiments reported in this paper were conducted in the Experimental Economics Laboratory at the University of New Mexico. Preliminary versions of this paper have been presented at the University of Mississippi, the Southern Economic Association Annual Meetings in November 1999, the University of Colorado at Boulder, Georgia State University, the University of Tennessee-Knoxville, and the Annual Meetings of the National Tax Association in November 2001.

* Corresponding author. Tel.: +1-404-651-0420; fax: +1-404-651-4985.

E-mail address: jalms@gsu.edu (J. Alm).

audit rule, in which each individual faces a fixed and predetermined probability of audit, regardless of his or her report. Many tax agencies do in fact randomly select some returns for audit. However, much audit selection is heavily dependent upon the information reported by taxpayers on their returns. Here the agency selects returns randomly for audit on the basis of information from the returns. With such an endogenous audit rule, the probability of audit is not fixed and constant but becomes variable and endogenous, depending in part upon the behavior of both the taxpayer and the tax agency.

There are many ways by which a tax agency can utilize information on the tax returns to determine whom to audit. One endogenous audit selection rule recognizes the dynamic aspect to compliance; that is, the tax agency may be able to use a taxpayer's history in targeting whom to audit, so that taxpayers known to have been non-compliant in the past will be audited more frequently in the future. Another endogenous audit rule applies this same approach to previous periods, so that individuals audited and found to be dishonest in the current period face the certain prospect that the tax agency will go back in time to previous period's declarations. Both of these rules have been shown to be more effective in deterring evasion than a simple random audit rule based only on current period declarations (Greenberg, 1984; Landsberger and Meilijson, 1982; Rickard et al., 1982; Alm et al., 1993).

Still another endogenous audit rule is based upon the deviations of each individual's tax report from the average of the reports of others in the taxpayer's cohort. For example, in the United States the Internal Revenue Service (IRS) uses the results of its previous experience with audited returns to devise a formula called the "discriminant index function" (DIF) that determines which tax returns to audit based upon items reported on the current returns. This formula estimates a "DIF score" for each return, with a higher DIF score indicative of a return with a higher likelihood of additional assessments in excess of audit costs. These DIF scores lie behind the so-called "audit flags": individuals who deviate from the average behavior of their assigned cohort send up a flag, and the IRS audits those for whom the flag indicates possible low compliance. The IRS does not reveal the details of this audit rule, but many tax professionals routinely publish guides that provide information about which deductions (and other reporting strategies) they believe are allowed and which they think are likely to be challenged.¹ In fact, IRS audits based upon the DIF score generate significantly more additional assessments than purely random audits, and roughly one-half of all audited returns in the United States are now selected by this approach (United States General Accounting Office, 1976, 1998, 1999). Many other countries follow a similar practice.

Tax professionals have long recognized the broad outlines of this audit process. Nearly every year, especially around April 15th, articles and books appear with titles such as "How to avoid a tax audit," most arguing that taxpayers should try to avoid waving a red flag in front of the IRS. For example, Aczel (1995) argues that large deductions relative to income increase the likelihood of an audit, with over 90 percent of all audits triggered by the size of deductions relative to income. He claims that his analysis shows that the audit probability goes up significantly when deductions are between 35 and 44 percent of adjusted gross income and that a ratio in excess of 44 percent is almost certain to trigger an audit. Other, more specific flags identified by tax professionals include larger than typical

¹ The IRS has consistently refused to disclose the specific components of the DIF formula and has even gone to court to maintain confidentiality.

medical deductions, mortgage interest deductions, travel and entertainment expenses, home office deductions, charitable donations, dependent exemptions, casualty losses, or tax shelter losses. Web sites are widely available that indicate the levels of itemized deductions or Schedule C business expenses relative to income that make a taxpayer “unlikely,” “likely,” or “almost certain” to be flagged for an audit.² The basic advice of these tax professionals is quite simple and easily summarized: “Don’t be different.”

This endogenous audit rule, therefore, implies that one’s chance of audit depends upon one’s behavior relative to all other taxpayers. Because the tax agency faces a budget constraint and cannot audit all returns, each individual can still cheat on the taxes, as long as he or she simply makes sure that his or her DIF score is lower than enough other taxpayers to fall below the agency’s threshold. With the agency essentially selecting a return for audit on the basis of the deviation between that return’s report and the average report of all others in the audit class, each individual must in principle guess the behavior of others when making his or her compliance decision. More realistically, given the millions of taxpayers who file returns, each taxpayer must guess the behavior of the “average taxpayer” upon whom his or her individual deviation is based.

This type of endogenous audit rule has not been examined in the tax compliance literature. However, work on coordination games is obviously relevant (Cooper et al., 1990, 1992; Van Huyck et al., 1990, 1997). Because audit selection is based upon relative reporting behavior, there is a game among the taxpayers, and taxpayers will have an incentive to coordinate their reporting behavior so that no taxpayer has a DIF score that deviates too much from the average. In this coordination game, there are multiple equilibria, and those equilibria that involve low reporting of tax liabilities are clearly preferred by the taxpayers as a group.³ However, these may not survive refinements (such as properness), since a taxpayer can reduce his or her audit probability by reporting a slightly higher tax liability than others in the audit class. The coordination equilibria that involve low reporting are, therefore, vulnerable to defection, at least for sufficiently high penalties.

The important issue here is whether this method of audit selection encourages or discourages tax compliance. More precisely, can taxpayers coordinate their compliance behavior when audits are based only on the DIF rule? Does application of the DIF rule with some tolerance in the allowed deviation affect compliance? Are there tactics available to taxpayers that allow them to achieve the zero-compliance equilibrium? If there are such tactics, can the tax authority devise methods that overcome them?

These are the issues examined in this paper. We report the results of laboratory experiments that investigate the compliance behavior of individuals when audits are conducted on the basis of deviations of reported return items among the taxpayers. Our experimental results indicate that individuals typically find it difficult to coordinate on the zero-compliance equilibrium, especially when the DIF audit rule is applied with some tolerance. However, pre-game communication (or “cheap talk”) that mimics the kinds of information-sharing provided by tax guides and professional opinions provides a mechanism that allows such coordination. Nevertheless, the tax authority is able to overcome this taxpayer coordination

² For other typical examples, see Daily (1998) and Block (1999).

³ The possibility of multiple tax reporting equilibria has been examined by Graetz et al. (1986).

by a subtle change in its audit selection rule, a change that does not increase the overall number of audits but that targets the audits in a slightly different way.

The next section presents the underlying theory of coordination games and relates this theory to the tax compliance game. The third section discusses our experimental design. Results are presented in the fourth section, and conclusions are in the final section.

2. Theory and hypotheses

In many games, there are multiple equilibria, and standard deductive equilibrium analysis is unable to determine which of these many possible equilibria the players will actually select. However, it is possible to argue that some Nash equilibria are implausible and will not be chosen. It is here where the notion of a “coordination game” enters. The essential feature of a coordination game is that there is no dominant strategy. Instead, there is a conditional best response that involves a matching of strategies across players.

The genesis of coordination games is the Schelling (1960) concept of a focal equilibrium, in which players in a game will have the highest payoffs if they can coordinate on a strategy. In the simplest version, the individuals are indifferent as to the actual strategy as long as all players choose the same one. The players have information that enables them to focus on a particular strategy that will improve the likelihood of reaching the coordinated outcome. In the context of a simple two-person coordination game, the payoffs to both players are higher when they can coordinate their strategies so as to lie on the diagonal in the payoff matrix.⁴

A similar game arises in the tax compliance setting. With compliance, there is a game among the taxpayers in a given cohort, instituted by the tax authority’s use of a conditional audit rule that selects returns for audit on the basis of some difference or deviation in reporting behavior among taxpayers. In the United States, this audit rule is the DIF rule.

The formulation of the DIF audit selection rule relies upon the Taxpayer Compliance Measurement Program (TCMP), which began in 1962 and was continued on a roughly 3-year cycle until 1988. The TCMP is a detailed line-by-line audit of a stratified random sample of approximately 50,000 individual tax returns. Each item on the return is examined by an auditor, who compares the amount actually reported by the taxpayer with the auditor’s estimate of the “true” amount. The taxpayer is liable for any additional tax payment plus interest and penalties uncovered by the audit. The IRS then uses statistical methods similar to discriminant analysis to estimate the relationship between the likelihood of a substantial audit assessment and individual tax return characteristics, with the returns separated by audit class; the dependent variable in this estimation is the auditor’s total recommended tax change for a return, and the independent variables are various return items reported by the taxpayer, entered in linear and non-linear forms. The output of this analysis is a formula that

⁴ An example from Van Huyck et al. (1990) illustrates a simple coordination game. Let (e_1, e_2, \dots, e_n) denote the actions of n players, where actions are limited to the set of integers with $0 < e_i < e^{\max}$. The payoff to each player π_i is defined as $\pi_i(e_i, \underline{e}_i) = a[\min(e_i, \underline{e}_i)] - be_i$, where $\underline{e}_i = \min(e_1, \dots, e_{i-1}, e_{i+1}, \dots, e_n)$, $a > b > 0$, and the players are assumed to have full information. The payoff-dominant strategy in this game requires each player to choose the maximum feasible action e^{\max} because $a > b$. However, any n -tuple of identical actions (e, e, \dots, e) is also a mutual best-response Nash equilibrium.

Table 1
Taxpayer A’s payoff space—DIF audit rule^a

Player A’s reported income	Player B’s reported income					
	5	4	3	2	1	0
5	3.50	3.50	3.50	3.50	3.50	3.50
4	3.20	3.80	3.80	3.80	3.80	3.80
3	2.90	2.90	4.10	4.10	4.10	4.10
2	2.60	2.60	2.60	4.40	4.40	4.40
1	2.30	2.30	2.30	2.30	4.70	4.70
0	2	2	2	2	2	5

^a The tax rate t equals 0.3, and the fine rate f is 1. Denoting true income as I and reported income of player A as R_A , player A’s payoff in the top right triangle and along the diagonal equals $[I - tR_A]$; the payoff in the bottom left triangle equals $[I - tR_A - ft(I - R_A)]$. Because the payoffs to player B are symmetric, the coordinated strategies are along the diagonal and are in bold; the coordinated strategies are also the Nash equilibria. With the addition of a random audit rule with a probability of audit p , the elements on the diagonal of this matrix equal $[I - tR_A - pft(I - R_A)]$, and are (from the upper left) 3.50, 3.68, 3.86, 4.04, 4.22, and 4.40; all other (off-diagonal) elements are unchanged.

assigns a “DIF score” to every tax return within the audit class, based upon reported return characteristics. The returns with the highest DIF scores are the ones predicted to have the greatest potential for a substantial audit assessment, and these returns are then selected for possible audit consideration, subject to the availability of examination resources.

This description makes it evident that a taxpayer’s probability of audit is based not only upon his or her reporting choices, but also upon these choices relative to other taxpayers in the cohort. In short, there is a taxpayer–taxpayer game that determines each individual’s chances of audit selection.

To illustrate, consider for simplicity a two-person game (the argument easily generalizes to a multi-person game), as well as a discrete setting in which each taxpayer has true income I of \$5 and must choose the amount of income to report, R . The individual pays taxes on reported income at the tax rate t , assumed to be 0.3 (or 30 percent). The individual pays no taxes on unreported income. However, if audited, then the individual is assumed to pay unpaid taxes plus an additional fine of unpaid taxes, so that the effective fine rate f is equal to 2. Now, if both individuals report \$5, then no one is audited, and each individual’s payoff is simply income less taxes, or \$3.50. Similarly, no one is audited if each reports \$4, \$3, and so on down to a common report of \$0. In these cases, each taxpayer again has the same payoff, equal to income less taxes on reported income. However, if one taxpayer reports, say, \$4 while the other reports \$5, then the taxpayer with the lower report will be audited, the unreported income will be detected (\$1), and the taxpayer must pay a fine of \$.60 (or $2 \times 0.3 \times \$1$). The payoff to this taxpayer is now only \$3.20 (or \$5 less \$1.20 taxes on reported income less \$.60 fine on unreported income). The payoff to the other taxpayer is \$3.50 (or \$5 less \$1.50 taxes on reported income).

The payoffs to the row player (player A) in the discrete strategy version of this game are shown in Table 1; the payoffs to the column player (player B) are symmetric. The coordinated strategies for players A and B are shown in bold type. There is no dominant strategy for either player, and any combination of equal reported incomes is a Nash equilibrium. Clearly, however, the coordinated strategies are Pareto-ranked by the taxpayers. The optimal

behavior for both players is to coordinate on reporting zero income because this yields the highest payoff: if the tax authority only audits on the basis of the difference, neither is audited and each succeeds in paying zero taxes.

More formally, the payoff to each player can be defined as follows. Denoting the reported income of player i as R_i , then player A's payoff π_A equals:

$$\pi_A(R_A, R_B) = \begin{cases} I - tR_A, & \text{if } R_A > R_B \\ I - tR_A, & \text{if } R_A = R_B \\ I - tR_A - ft(I - R_A), & \text{if } R_A < R_B \end{cases}$$

where each player receives the same true income I . The payoffs to player B are symmetric. The coordinated strategies are equal levels of unreported income, and the payoff-dominant strategy is to report zero income.

However, it is well known that coordination games provide special difficulties for the selection of equilibrium strategies (Sefton, 1999). Consequently, an issue in coordination games is the equilibrium selection. In games such as those shown in Table 1, the players can coordinate on many strategies; in fact, there is an infinite number if the strategy space is made continuous or if mixed strategies are played. However, there are some strategies that the players will prefer, and the behavioral question has been whether they will be able to select these. The presence of multiple equilibria that are robust to conventional refinements makes it difficult to predict play in such games absent some additional factors that may make some outcomes more focal than others. The presence of millions of taxpayers also makes it difficult to predict play.

It is, therefore, an empirical question whether the taxpayers in a given cohort can coordinate on the zero-compliance outcome. This is the first issue addressed here:

1. *Can taxpayers coordinate on the zero-compliance Nash equilibrium when audits are based solely on the DIF rule?*

Experimental evidence (Cooper et al., 1990; Van Huyck et al., 1990, 1997) suggests that taxpayers will find it difficult to achieve the payoff-dominant outcome, even when the number of players is small.

Due to an unavoidable imprecision in audit procedures, the tax authority is typically not able to determine non-compliance perfectly with its DIF rule. A DIF audit rule may, therefore, result in the targeting of taxpayers found to be compliant if it is applied too stringently. Consequently, the tax authority may choose to apply the flags with a tolerance factor, especially since the tax authority is also generally subject to budget limitations; that is, the authority will not audit a return unless the deviation from the average exceeds some minimum cutoff. If no one reports a deviation greater than the cutoff, then no one is audited; if the returns exhibit deviations greater than the cutoff, then the return with the greatest deviation is selected for audit. A second issue examined is

2. *Can taxpayers coordinate on the zero-compliance Nash equilibrium when audits are based solely on the DIF rule, applied with a cutoff?*

It seems likely that the presence of a cutoff will create some uncertainty on the part of the taxpayers, making coordination on the zero-compliance outcome even more dif-

ficult. Put differently, a minimum cutoff makes the level of required coordination less precise, thereby obscuring any potential focal equilibrium of the taxpayers and increasing compliance.

Various factors may facilitate coordination on certain equilibria. In particular, shared information regarding tax deductions or other reporting behavior may enable taxpayers to coordinate behavior, leading to lower compliance levels. Many tax professionals provide such information, and trade and professional groups often provide tax guides describing classes of deductions that can be “safely” taken and income computation procedures that are “generally” accepted by the tax authority associations; the IRS itself releases information on the average amounts of reported return items by income class. All this information constitutes a form of “cheap talk” that may facilitate coordination on a reporting strategy for the cohort (Cooper et al., 1992). The third issue examined is, therefore,

3. *Can information-sharing via taxpayer communication help taxpayers coordinate on the zero-compliance Nash equilibrium when audits are based solely on the DIF rule?*

In the context of the experiments here, this “cheap talk” is provided by pre-play communication.

If the cohort is successful in coordinating on the zero-compliance equilibrium, the tax agency may be able to overcome this outcome by adding an additional audit rule to the DIF rule. According to the [United States Department of the Treasury, Internal Revenue Service \(2000\)](#) and the [United States General Accounting Office \(1998, 1999\)](#), the IRS uses about 40 audit programs to select returns for audits, and most of these employ non-DIF techniques to select individual returns for audit. A return may be selected when third-party sources of information, such as Forms 1099 and W-2, do not match the information reported on the tax return, or when information on potential non-compliance is received from other sources, such as other agencies, the media, public records, or informants. A return may be selected to examine the behavior of taxpayers in a particular “market segment” (e.g. farmers or waiters) or to examine a specific, questionable item. A return may also be targeted on the basis of its preparation by a tax preparer already under IRS scrutiny. In total, these other methods account for roughly one-half of all IRS audits.

Nearly all of the IRS audit programs now rely upon non-random audit selection of returns for audit. Traditionally, the only IRS program that used random selection was the TCMP, and this program has not been in operation since 1988. Currently, the IRS has several audit programs targeted at specific classifications, within which returns are then selected randomly for audit.⁵ However, very few IRS audits are now based on purely random audit selection, mainly because the IRS has found that non-random selection of returns yields higher returns than random selection.

⁵ According to the [United States General Accounting Office \(1998\)](#), these projects focus upon taxpayers claiming the Earned Income Tax Credit (EITC), claiming dependent exemptions duplicated on other returns in Florida and in other specific locations, operating certain kinds of eating and drinking establishments in Ohio, filing questionable Schedule Cs in Illinois, claiming false business losses to be eligible to claim the EITC in Georgia, and not paying the self-employment tax in Missouri. The number of taxpayers in these target groups ranged from 2348 to 15 million. The overall number of audits of these target groups was less than 0.05 percent of all IRS audits for the years 1994–1996.

In this overall context, the simplest additional audit rule that can be implemented, and implemented in a way consistent with the IRS use of conditional audit programs (and IRS budget constraints), is a conditional random audit rule. A conditional random audit is implemented as follows: if both players report the same amount, then one player will be selected with a fixed and predetermined probability for audit. Of course, the tax agency can no doubt target its audit selection more precisely than implied by such a simple audit rule. However, this greater accuracy comes at greater cost. Consequently, it seems useful to examine the effects of a conditional random audit rule on the outcome of the earlier tax compliance game.

The expected payoffs to player A for this modified game are reported as a note to [Table 1](#), where the probability of audit p under the random rule is 0.2 (or 20 percent). Consider the joint strategies of reporting zero income. The expected payoff to A is now \$4.40, not \$5; there is a 0.2 probability of being audited and having to pay a penalty of \$3, so the expected penalty equals \$.60. Note, however, that if A reports income of one dollar when B reports zero, the payoff to A is \$4.70. Consequently, the best response for A is no longer to report zero income when the random audit is in effect. The theory would, therefore, support the joint use of a DIF rule and a non-DIF rule based upon a conditional random audit. A fourth issue examined here becomes

4. *Can the tax authority increase compliance by combining a DIF audit rule and a random audit rule?*

The next section presents the experimental design used to investigate these issues.

3. Experimental design

The experimental design captures the essential features of the voluntary income reporting and tax assessment system used in many countries.⁶ Human subjects in a controlled laboratory environment receive income, and each must decide how much of this income to report. Taxes are paid on reported income, but not on unreported income. However, unreported income may be discovered with some probability, and the subject must then pay a fine on the unpaid taxes. This reporting, audit, and penalty process is repeated for a given number of rounds that represent tax periods and is replicated with different sets of subjects. At the completion of the experiment, each subject is paid an amount that depends upon his or her performance during the experiment.

Subjects are recruited from undergraduate classes in economics and business, and they are organized into groups of five persons with multiple groups in each session. They are told that the experiment will last an unknown but predetermined number of periods called rounds; the actual number of rounds is 25. The currency used in the experiment is called tokens, and subjects are told that all tokens that they earn during the experiment will be redeemed for cash at the end of the experiment at a fixed conversion rate of six tokens per dollar.

At the beginning of each round, subjects are randomly assigned an amount of income, denoted in tokens and drawn from a uniform and discrete distribution with the support

⁶ The full set of experimental instructions is available upon request.

2.0 to 4.0 tokens, in increments of 0.2 tokens. Within a group, all subjects are assigned the same level of income in a given round; this amount varies from round to round, and the equal assignment is common knowledge. Each subject chooses a level of income to report, with taxes paid at the rate of 30 percent on reported income; the tax rate is the same in all rounds and all treatments. After all subjects have disclosed their incomes and paid their taxes, a subject may be selected for an audit, in which any unpaid taxes for the current round will be discovered and the subject must pay the unpaid taxes plus a penalty equal to unpaid taxes; the effective fine rate is, therefore, two and remains constant in all rounds and treatments. The procedures for selecting subjects for audit are described later.

Subjects have no previous experience with this experimental setting, and they are not allowed to communicate with one another during the session; as discussed later, in some treatments subjects are permitted to discuss strategies before the experiment begins. All sessions begin with the instructions being read aloud while the subjects follow along with their own copies, a procedure that ensures common knowledge. In keeping with previous research (Alm et al., 1990, 1992a,b), the instructions use neutral terminology (e.g. “check” versus “audit”) to avoid context or framing effects that may bias subject choices unpredictably; tax terminology is used here to facilitate the discussion. Three practice rounds are given, and procedural questions are answered. The full experiment then begins. Each session lasts slightly under one hour, and each subject earns between \$10 and \$14, depending upon his or her performance during the experiment. Certain parameters (e.g. the tax rate and the penalty rate) are fixed throughout the experiments. All audits investigate only the current period disclosure.

Several audit selection rules are examined. The basic audit rule is a stylized version that reflects the DIF approach used by the IRS. For each subject the deviation between his or her reported income and the average reported income for the group is calculated. The subjects are then ranked on the basis of the deviations, and the highest deviation in each group is selected for audit. Of course, the actual DIF approach by the IRS is based upon a large array of items reported on an individual’s tax return, not simply one item. Nevertheless, the essence of the DIF approach is to select returns based upon the deviation of a taxpayer’s report from that of his or her cohort, and our stylized version captures this critical element. Also, our basic audit rule reflects the limited audit resources of the tax agency because our rule requires the selection of only a single return.

The basic audit rule examines the return with the highest deviation, regardless of the size of the deviation. This rule is modified in some treatments by allowing some tolerance, or cutoff, in the allowed deviation. If no one reports a deviation greater than the cutoff, then no one is audited; if more than one subject reports an amount of income that gives a deviation in excess of the cutoff, then the subject with the largest deviation is audited. The introduction of a minimum cutoff is intended to capture the imprecision in a tax agency’s audit procedures as well as the agency’s budget constraint. The tax authority is not able to determine non-compliance perfectly on the basis of the DIF function, and the authority can set a cutoff for the deviation that will trigger an audit in order to avoid largely unproductive audits. The level of the minimum cutoff is varied, from 0 (the basic audit rule) to 0.25 and to 0.50 tokens. Variation in the minimum cutoff also allows investigation of the difficulty in coordinating reporting strategies. It is expected that a

Table 2
Experimental design and results^{a,b}

	Minimum differential cutoff		
	0.0 tokens	0.25 tokens	0.5 tokens
DIF audit rule with cheap talk (CT)	Treatment 1 (25) 0.550 (0.058)	Treatment 2 (10) 0.631 (0.072)	Treatment 3 (15) 0.748 (0.044)
DIF audit rule with CT	Treatment 4 (10) 0.017 (0.025)		
DIF + random audit rule with CT	Treatment 5 (10) 0.240 (0.082)		
DIF + random audit rule without CT	Treatment 6 (10) 0.705 (0.113)		

^a All treatments last 25 rounds. In all treatments, the tax rate is 0.30, the fine rate is 2, subjects are organized into groups of five persons, and all subjects in a group receive the same level of income in a given round, an amount that varies from round to round. The number of subjects is shown in parentheses in the upper portion of each cell.

^b The “average compliance rate” is shown in the lower box of each treatment. It is calculated as the average of the individual compliance rates over all subjects, rounds, and groups in a given treatment, where the individual compliance rate is calculated as the ratio of reported income to true income. The number in parentheses is the standard deviation of the individual compliance rate.

greater cutoff will make it more difficult for subjects to coordinate on the low compliance equilibria.⁷

In some treatments, one additional audit rule is introduced. If no one in a given group exceeds the minimum cutoff deviation, a random audit is conducted in which one person in the group of five is randomly selected for audit with a fixed and predetermined audit probability. This additional audit rule is meant to capture the IRS use of non-DIF methods of audit. As discussed earlier, the IRS uses a variety of non-DIF techniques to select individual returns for audit. The use of a conditional random audit rule is perhaps the simplest way of capturing experimentally the IRS practice of supplementing its DIF audit rule with other conditional audit rules, within an overall budget constraint that limits the total number of audits.

A further treatment is conducted in which subjects are permitted to discuss their strategies among themselves prior to beginning the experiment but after the instructions have been read and the practice rounds conducted. This “cheap talk” mimics the provision of information within a cohort via tax professionals or trade publications. The cheap talk phase consists of a 5-min period following the practice rounds during which the subjects are free to discuss the experiment among themselves. They are monitored to ensure that no threats or agreements outside the scope of the experimental setting are made. Subjects are assigned to different stations after the cheap talk phase to reinforce the fact that they cannot use the practice rounds to determine their group assignment.

The experimental design is summarized in the top portion of each cell of Table 2. Treatment 1 examines the ability of the subjects to coordinate their strategies with the basic DIF audit rule with a zero cutoff (Issue 1). Note that coordination should be more easily achieved

⁷ In this regard, Alm et al. (1992a) report experimental evidence in which greater taxpayer uncertainty about audit procedures often leads to greater tax compliance. See also the theoretical analysis of Scotchmer and Slemrod (1989), who reach a similar conclusion.

in a group of five than in larger groups; if we do not observe coordination on the zero-tax equilibrium in a small group of five, then it is likely that larger groups of taxpayers will also find it difficult to achieve such coordination. Treatments 2 and 3 examine the impact of a greater tolerance in the cutoff on coordination (Issue 2). Treatment 4 investigates the role of cheap talk (CT) (Issue 3), and Treatments 5 and 6 add a random audit rule to the DIF rule, with and without the cheap talk (Issue 4).

4. Experimental results and discussion

The average compliance rates across groups by treatment are presented in the bottom portion of each cell of Table 2, where the average compliance rate is calculated as the ratio of reported income to true income, averaged across all subjects, rounds, and groups in a given treatment. Individuals are clearly unable to coordinate on the zero-compliance, payoff-dominant Nash equilibrium (Issue 1), as shown by the average compliance rate of 55.0 percent in Treatment 1. In fact, the average compliance rate in any particular round only once falls below 20 percent (Group 1), and in one group (Group 2) the compliance rate averages over 90 percent for all rounds (Fig. 1); for Groups 1, 4, and 5, the compliance rate tends to decline over time, but compliance rises over the rounds of the experiment for Groups 2 and 3. These results are consistent with those of Cooper et al. (1990) and Van Huyck et al. (1990, 1997), who found that subjects were typically unable to coordinate on the payoff-dominant Nash equilibrium.

As expected, allowing some tolerance in the cutoff makes coordination on the zero-compliance equilibrium even more difficult (Issue 2). The average compliance rate increases from 55.0 percent in Treatment 1 to 63.1 percent in Treatment 2 (the 0.25 differential cutoff rule). Compliance increases still further in Treatment 3 (the 0.50 cutoff), to nearly 75 percent. The differences between these average compliance rates are all statistically significant.

However, pre-play communication reverses the coordination failure (Issue 3), as shown in Table 2 and Fig. 2. In the presence of cheap talk in Treatment 4, the average compliance rate falls virtually to zero (or 1.7 percent). One subject in each of the two “cheap talk” (CT) groups was able to convince the others that reporting zero income is the optimal

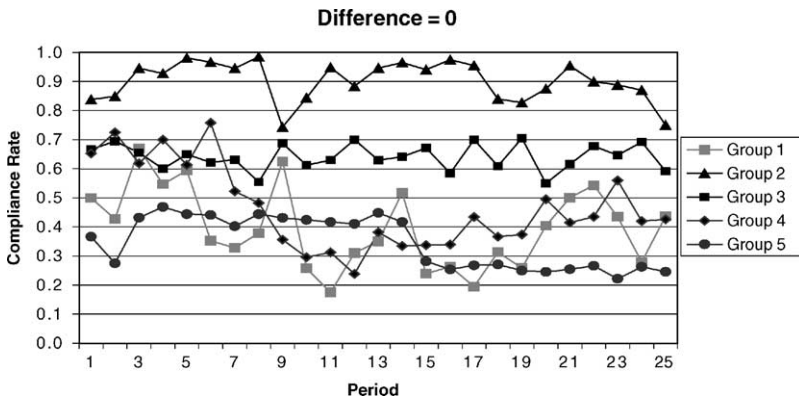


Fig. 1. Treatment 1 (DIF audit rule with 0.0 minimum differential cutoff): group results.

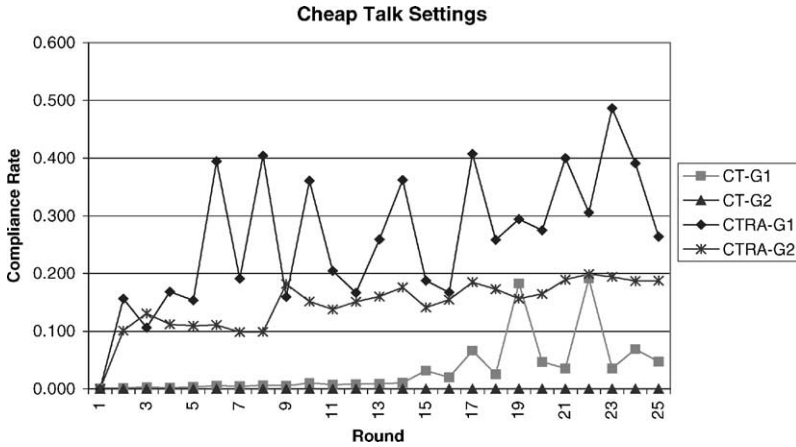


Fig. 2. Treatments 4 and 5 (CT with and without random audits): group results.^a
^a “CT” refers to the Treatment 4 sessions in which pre-play “CT” communication is allowed. “CTRA” refers to the Treatment 5 sessions in which pre-play “CT” communication is allowed with a “Random Audit” process. “G1” and “G2” denote the specific “Group” in each treatment.

strategy, arguing that “No one will be audited, and we will take home all of our earnings.” Group 2 subjects in this CT treatment coordinated immediately upon the zero-compliance equilibrium and remained there for the entire experiment; Group 1 subjects were slightly less successful. Still, the average compliance rate is only 3.3 percent.⁸

It is striking that the simple addition in Treatment 5 of a conditional “random audit” component to “cheap talk” (CTRA) is largely able to mitigate the effects of pre-play communication, or Issue 4. The initial discussion among the subjects in Treatment 5 focused on the zero-compliance coordinated strategy, just as in the earlier cheap-talk treatment. The individuals recognized that one person would be audited each round, but they argued that overall they would still be better off if they allowed this to happen randomly throughout the session. As shown in Fig. 2 and Table 2, in both CTRA groups the compliance rate was uniformly zero in the first round, and a randomly selected individual was audited in each group. However, the subjects became increasingly unwilling to expose themselves to the risk of a random audit. The compliance rate increased immediately and tended to rise over time; after round 1, no further random audits were conducted because the subjects did not coordinate on the zero-compliance outcome.⁹

⁸ In Group 1, the subject who made this argument was especially articulate and was wearing a reserve officer training corps (ROTC) uniform. The other subjects agreed with his assessment, and the discussion ended quickly. During this session, the ROTC student started the experiment reporting 0.01 tokens; because this amount is not different than 0 at two decimal places, no one in the group was audited. As the session progressed, this student increased his reporting level until an audit was triggered. The other subjects attempted to restore the zero-compliance equilibrium, but could not do so, so that the average compliance rate tended to rise slightly over time. The ROTC subject was never audited.

⁹ Note also that adding a conditional random audit mechanism to the cutoff rule increases compliance relative to the cutoff rule alone (Treatment 6 versus Treatment 1). The average compliance rate increases from 55.0 to 70.5 percent with the addition of the random audit rule, and this occurs with no additional audits.

Table 3
Random effects estimation (GLS)—results from individual data^a

Independent variable	Dependent variable: compliance rate			
	Model 1	Model 2	Model 3	Model 4
Constant	0.550 (13.33)	0.573 (13.58)	0.563 (16.38)	0.587 (21.17)
Wealth		−0.0007 (3.05)		−0.0007 (3.04)
Audit previous round			−0.068 (5.50)	−0.074 (5.89)
DIF audit rule (0.25 cutoff)	0.081 (1.05)	0.082 (1.05)	0.082 (1.29)	0.083 (1.67)
DIF audit rule (0.5 cutoff)	0.199 (2.95)	0.197 (2.91)	0.195 (3.48)	0.193 (4.45)
DIF audit rule with CT	−0.533 (6.91)	−0.529 (6.82)	−0.542 (8.45)	−0.539 (10.86)
DIF + random audit rule with CT	−0.346 (4.49)	−0.344 (4.43)	−0.346 (5.40)	−0.344 (6.93)
DIF + random audit rule without CT	0.155 (2.01)	0.154 (1.99)	0.155 (2.42)	0.154 (3.10)
Wald Chi-square	112.68	120.63	193.27	315.95
R ² within	0.000	0.005	0.012	0.017
R ² between	0.604	0.604	0.630	0.633
R ² overall	0.459	0.461	0.482	0.485
Rho ^b	0.559	0.564	0.463	0.326
Observations	2000	2000	2000	2000

^a z-statistics are in parentheses.

^b Rho is the fraction of the variance due to the random effects.

Table 3 presents the results of econometric estimation of individual behavior, using the individual round-by-round data in a set of random effects models.¹⁰ Four model specifications are reported, based upon different inclusion of explanatory variables that are potentially endogenous. The dependent variable in each model is the individual's compliance rate in a given round, defined as the ratio of reported income to true income in that round (compliance rate).¹¹ The independent variables include session-specific variables for each treatment, with the omitted category the DIF audit rule with a 0.0 cutoff (or Treatment 1). The independent variables also include in Models 2, 3, and 4 individual-specific information: the individual's accumulated earnings (wealth) and a dummy variable equal to 1 if the individual was audited in the previous round and 0 otherwise (audit previous round). These variables are potentially endogenous; however, a Hausman specification test rejects the hypothesis that they are in fact endogenous, so that our preferred model is Model 4.

With the individual effects included, the results for the audit treatment variables strongly reinforce our earlier discussion. In particular, the coefficients on the audit treatments are generally significant, and their magnitudes are quite robust across the three models. For example, compliance is higher under the 0.25 and 0.5 DIF rules than when the cutoff is zero. Cheap talk always reduces compliance relative to the zero cutoff treatment; however, the addition of a random audit rule to the cheap talk treatment significantly reduces the

¹⁰ We are grateful to David Grether for the suggestion that we employ a random effects model.

¹¹ The dependent variable can also be redefined in terms of the ratio of unreported income to true income or the level of unreported income. The results are unaffected.

effect of cheap talk, and compliance increases relative to the cheap talk treatment without the random audit rule. Also, in the absence of any cheap talk, compliance is higher when the random audit rule is used together with the DIF audit rule.

Interestingly, compliance is negatively correlated with an individual being audited in the most recent period. If the audits were based only upon a random process, this result might be interpreted as a “gambler’s fallacy” effect. However, in the presence of an endogenous audit rule, such behavior is a best response strategy if the individual expects others to lower their compliance because they were not audited in the previous round. Also, a significant coefficient on wealth provides a test of the ability of subjects to coordinate their reporting behavior. If subjects are able to coordinate, accumulated earnings should not affect compliance, although there could be a wealth effect that captures attitudes toward risk. The results in [Table 3](#) are consistent with the inability of subjects to coordinate, at least in the absence of cheap talk.

In sum, it is clear that subjects are largely unable to coordinate on the low (or zero) compliance equilibrium in those treatments without cheap talk (Treatments 1 and 6). It is equally clear that cheap talk with no random audit component (Treatment 4) facilitates coordination at very low levels of compliance. As emphasized earlier, the addition of a random audit rule to the cheap talk session (Treatment 5) is able to mitigate this coordination to a large extent, but even so the opportunity to engage in cheap talk still permits some coordination at low compliance levels.¹²

5. Conclusions

Does the IRS use of a DIF audit rule generate higher or lower levels of tax compliance? The experiments reported here suggest that a DIF rule is often able to achieve high levels of compliance; that is, taxpayers find it difficult to coordinate their reporting behavior on the zero-compliance outcome. However, a certain amount of luck may be involved, and evidence from the cheap talk session is that information-sharing may be very successful in aiding the taxpayers achieve an extremely low compliance rate. In the face of this communication, the use of an additional audit rule is very useful to the tax authority, generating compliance significantly above the level achieved by a conditional audit rule alone. Importantly, this higher compliance is achieved without the expenditure of greater audit resources. The actual audit rate is virtually identical across all treatments, at 20 percent in each round,¹³ and, even if audit rates could be increased at low cost to the tax agency, there is much experimental evidence that the increases in compliance that result become significantly smaller as the probability of detection increases.¹⁴ In short, as long as taxpayers are unable

¹² This result holds even when only the later rounds of the sessions are examined.

¹³ The audit rate is slightly lower in Treatments 2 and 3, given the tolerance in the minimum differential cutoff, and the audit rate is zero in those rounds of Treatment 4 in which subjects coordinate perfectly on the zero-compliance equilibrium.

¹⁴ For example, in an identical experimental design, [Alm et al. \(1993\)](#) find that increases in a purely random audit rate in increments from 5 to 50 percent are only able to increase the average compliance rate from 28 to 49 percent. Similarly, [Alm et al. \(1992b\)](#) find diminishing impacts from increases in audit rates.

to coordinate perfectly on the zero-compliance equilibrium, a DIF audit rule in combination with a conditional random audit rule generates no more audits than a DIF rule by itself.

In fact, the IRS splits its audit rules roughly equally between DIF and non-DIF methods. One may perhaps think of the IRS audit strategy as one in which the tax authority has a fixed budget and employs the following audit scheme: conduct DIF audits first, then conduct conditional random audits if the DIF audits do not exhaust the audit budget, and also conduct a small subset of purely random audits to generate the data needed to construct the DIF audit rule (and also to generate some taxpayer uncertainty).¹⁵ However, this audit strategy was determined, our results suggest that it is a policy that is well-grounded, at least for a tax agency interested in maximizing its net revenues. In a world in which taxpayers are informed about IRS use of “audit flags,” the combination of DIF audits and non-DIF audits is able to generate substantially more compliance than DIF audits alone, even when the non-DIF method is based on a simple audit selection rule, and these additional revenues are collected with virtually no impact on audit costs.

Acknowledgements

We thank Brian Erard, Jonathan Hamilton, William Shuggart, Arthur Snow, Robert Tollison, and an anonymous referee for helpful comments. We are especially grateful to David Grether for comments that led to substantial improvements in our empirical analysis.

References

- Acelz, A.R., 1995. *How to Beat the I.R.S. at its Own Game: Strategies to Avoid—and Fight—an Audit*. Four Walls Eight Windows, (revised edition). New York, NY.
- Alm, J., Cronshaw, M.B., McKee, M., 1993. Tax compliance with endogenous audit selection rules. *Kyklos* 46, 27–45.
- Alm, J., Jackson, B.R., McKee, M., 1992a. Institutional uncertainty and taxpayer compliance. *American Economic Review* 82, 1018–1026.
- Alm, J., McClelland, G.H., Schulze, W.D., 1992b. Why do people pay taxes? *Journal of Public Economics* 48, 21–38.
- Alm, J., McKee, M., Beck, W., 1990. Amazing grace: tax amnesties and compliance. *National Tax Journal* 63, 23–38.
- Block, J., 1999. *Julian Block’s Tax Avoidance Secrets*. Larchmont, New York, NY.
- Cooper, R.W., DeJong, D.V., Forsythe, R., Ross, T.W., 1990. Selection criteria in coordination games: some experimental results. *American Economic Review* 80, 218–233.
- Cooper, R.W., DeJong, D.V., Forsythe, R., Ross, T.W., 1992. Communication in coordination games. *Quarterly Journal of Economics* 53, 739–771.
- Daily, F.W., 1998. *Stand Up to the IRS*, fourth ed. Nolo Press, Berkeley, CA.
- Graetz, M., Reinganum, J.F., Wilde, L.L., 1986. The tax compliance game: toward an interactive theory of tax enforcement. *Journal of Law, Economics, and Organization* 2, 1–32.
- Greenberg, J., 1984. Avoiding tax avoidance: a (repeated) game-theoretic approach. *Journal of Economic Theory* 32, 1–13.

¹⁵ In this regard, the inability of the IRS to conduct random audits for its TCMP program since 1988 represents a significant limitation on its enforcement activities.

- Landsberger, M., Meilijson, I., 1982. Incentive generating state dependent penalty system. *Journal of Public Economics* 19, 333–352.
- Rickard, J.A., Russell, A.M., Howroyd, T.D., 1982. A tax evasion model with allowance for retroactive penalties. *Economic Record* 58, 379–385.
- Schelling, T.C., 1960. *The Strategy of Conflict*. Harvard University Press, Cambridge, MA.
- Scotchmer, S., Slemrod, J., 1989. Randomness in tax enforcement. *Journal of Public Economics* 38, 17–32.
- Sefton, M., 1999. A model of behavior in coordination game experiments. *Experimental Economics* 2, 151–164.
- United States Department of the Treasury, Internal Revenue Service, 2000. *Examination of Returns, Appeal Rights, and Claims for Refund*. Publication 556 (Revised, November 2000). Government Printing Office, Washington, DC.
- United States General Accounting Office, 1976. *How the Internal Revenue Service Selects Individual Income Tax Returns for Audit*. Report to the Joint Committee on Internal Revenue Taxation, Congress of the United States. Government Printing Office, Washington, DC.
- United States General Accounting Office, 1998. *IRS' Use of Random Selection in Choosing Tax Returns for Audit*. Report to the Honorable Paul Coverdell, US Senate, GAO/GGD-98-40. Government Printing Office, Washington, DC.
- United States General Accounting Office, 1999. *IRS' Return Selection Process*. Letter Report, GAO/GGD-99-30. Government Printing Office, Washington, DC.
- Van Huyck, J.B., Battalio, R.C., Beil, R., 1990. Tacit coordination games, strategic uncertainty, and coordination failures. *American Economic Review* 80, 234–248.
- Van Huyck, J.B., Battalio, R.C., Rankin, F.W., 1997. On the origin of convention: evidence from coordination games. *Economic Journal* 107, 576–596.