Behavioral Experiments of Alternative Reporting Regimes: Transparency vs. Burden

Laura Kalambokidis, University of Minnesota; Alex Turk, Internal Revenue Service; and Marsha Blumenthal, University of St. Thomas

Among the puzzles the Internal Revenue Service (IRS) faces in enforcing the U.S. income tax is how to stretch its limited resources so as to maximize taxpayer compliance. Random enforcement would be optimal only under a set of restrictive assumptions: (a) all taxpayers exhibit the same degree of noncompliance, paying the same share of their true tax liabilities, (b) the IRS is able to detect all noncompliance during an audit, and (c) audit costs are also proportional to true tax liabilities. But, in the first place, it is unlikely that all taxpayers are equally noncompliant. Instead, they differ in many ways that can affect compliance behavior: their occupations, the sources of their incomes and the various consumption choices they make, to name a few. These differences grant some taxpayers opportunities that others do not have to understate their liability—or more opportunities to do so than other taxpayers. Moreover, even if it were possible to eliminate these opportunities, people might still be endowed with individual differences in their underlying propensities to comply with tax laws.

At least partly in response to these differences, the IRS currently expends some resources to identify which taxpayers are to receive enforcement efforts. Alternatively, if taxpayers could be induced to reveal or signal their compliance propensities, a portion of those resources could be redirected elsewhere. How might we design a revelation strategy? One possibility that we study here is to offer people a choice of taxpaying regimes, coupled with adjustments in their taxpaying burden. Beyond self-revelation, this scheme also enables us to explore whether compliance changes. The idea is that a compliant taxpayer with a low propensity to cheat might prefer a regime that eliminated any cheating opportunities while also offering a reduced compliance burden. On the other hand, a noncompliant taxpayer with a higher propensity to cheat might prefer a regime that did offer opportunities to cheat, willingly accepting a higher compliance burden.

Opportunities to cheat are reduced when a taxpayer's income or deductions are transparent to the tax administrator, for example because they are required to be reported to the administrator by the source. In some cases, tax administrators allow taxpayers to trade transparency—and with it the chance to underreport income or overstate deductions—for reduced compliance burden. For example, the IRS maintains voluntary compliance agreements for industries where tipping of service-providers is customary. Under these programs, employers and employees may benefit from reduced compliance burdens and/or reduced threat of IRS examinations.[1] In exchange, they accept restrictions on their ability to underreport taxable tip income. As another example, under the Compliance Assurance Program, certain corporate taxpayers work with the IRS to identify and resolve potential compliance issues prior to filing a return. In return for being transparent and disclosing all materials to the IRS, the taxpayer receives certainty regarding their tax liability and potentially reduces their own resources dedicated to the examination process.

This paper presents the results of a laboratory experiment designed to explore three main questions: (1) whether offering a regime choice can serve as a sorting mechanism, separating (relatively) compliant from noncompliant taxpayers; (2) how much taxpayers are willing to pay for a reduction in burden, conditioned on their income; and (3) whether the compliance behavior of taxpayers who are able to choose a reporting regime changes in ways that improve the overall level of compliance.

We constructed three experimental treatments. In each treatment, subjects earned income by completing tasks on a computer terminal, and a tax was levied on income. In the first treatment, subjects reported their earnings to the experiment authority, and the reports were randomly audited, with penalties for underreporting. In this simplified experimental setting, differences in reporting compliance reflect only differing propensities to cheat, as everyone has the same occupation, the same income source, and there are no tax-sensitive consumption opportunities. The results of this treatment allow us to identify the subjects' propensities to underreport earnings and how that propensity varies with earnings and the tax, audit, and penalty rates. Among noncompliant subjects, we also learn how the amount of understatement varies with earnings and the experiment parameters.

The second treatment introduced a reporting burden in the form of an earnings worksheet—requiring the subject to perform a calculation—plus an on-screen form in which to enter earnings and several other pieces of information. Subjects could choose to pay a fee to be relieved of the reporting burden. If the subject paid the fee, the computer automatically and instantaneously recorded their earnings. Varying the fee, we trace out the schedule of subjects' willingness to pay for (demand) burden reduction. We also construct individual willing-to-pay indices for each subject.

Treatment 3 presented subjects with a choice of two tax reporting regimes: a transparent, variable burden regime and a non-transparent, high burden regime. The results of this treatment allow us to identify how well the propensity to cheat, as identified in Treatment 1, and willingness to pay for burden reduction, as identified in Treatment 2, explain regime choice. As we also vary the reporting burden, we study how the quality of separation and the overall level of compliance vary with burden.

In the following sections, we review the salient literature, describe the experiment, describe subject recruitment and characteristics, present results, and offer some conclusions.

Literature Review

Theoretical analysis of tax compliance behavior began in the deterrence models of Allingham and Sandmo (1972) and Yitzhaki (1974), following the earlier work of Becker (1968). For the United States, where audit and penalty rates are comparatively low, these models predict substantially more tax evasion than have usually been observed. In a more recent paper, Mark Phillips (2011) argues that replacing constant deterrence parameters with the IRS's targeted and endogenously determined audit and detection probabilities, compliance behavior does respond in the direction deterrence theory would predict. That is, for example, the 2007 net misreporting percentage for income subject to complete information reporting (wages and salaries) was only 1.2 percent while that for income not subject to any information reporting was 53.9 percent. To augment the deterrence theory, researchers have sought to model alternative explanations for compliance behavior. Some work focuses on the influences of social norms (Posner, 2000) or of how the government treats taxpayers (Feld & Frey, 2002). Other research enlarges the compliance paradigm beyond deterrence to include the provision of information and services (Andreoni et al., 1998, Slemrod et al., 2001, Alm, Jones, Cherry, and McKee, 2011).

While we know that individual taxpayers differ in important ways and that a tax authority often uses audit selection criteria to target and exploit such differences, comparatively few contributions to this literature address these differences. Alm, Cronshaw and McKee (1993) show experimentally that audit selection using endogenous rules results in enhanced compliance. Taxpayer differences may also be identified via self-selection mechanisms. For example, among business taxpayers, Chu (1990) proposes a separation into alternative plea-bargaining regimes on the basis of firm profit levels, while Cowell (1990) suggests that when multinational firms transfer income between sub-units, they are in effect self-selecting between alternative national tax regimes. For individuals, Falkinger and Walther (1991) allow taxpayers to choose between two tax regimes, with the tax authority subsequently using the separation to target its audits. Theirs is an Allingham/Sandmo/Yitzhaki model in which there are two types of taxpayers: "big fish" who evade a lot and "small fish" who evade only a little. To identify or separate the types, taxpayers are offered a choice between a BONUS and a NO BONUS regime. In the BONUS regime, likely attractive to "small fish," there is a lower tax rate (the bonus) and a higher penalty for evasion. In the NO BONUS regime, designed to attract "big fish," there is no lower tax rate but the penalty rate for evasion is lower. The tax authority sets the bonus tax rate reduction and the penalty rate increase so that "small fish" are better off in the BONUS regime and "big fish" in the NO BONUS regime. They show that when the tax authority subsequently raises the audit probability for those choosing NO BONUS, taxpayers who remain in the regime evade less, paying more tax. On the other hand, those "big fish" who migrate to the BONUS regime also evade less and pay more tax. While it is possible that some "small fish" will evade more (and pay less tax), the tax authority could eliminate this by setting the BONUS penalty rate sufficiently high. The end result here is that the small fish experience a welfare gain (lower tax rate), the big fish are induced to reduce their evasion, and the government receives higher tax yields.

Raskolnikov (2009) supposes two types of taxpayers, gamers and non-gamers. The gamers are either aggressive tax evaders whose behavior is clearly not justifiable, or aggressive tax avoiders, whose behavior can be justified by ambiguities in tax law. Non-gamers are taxpayers motivated other than by expected tax penalties, for example, by habit, a sense of duty, a desire to avoid guilt or shame, or a wish to behave as others do. Like Falkinger and Walther, he proposes that the IRS ask all taxpayers to choose between two tax regimes, here a compliance regime (CR) and a deterrence regime (DR). The CR features a lower penalty rate, an increased probability of conviction, and a few special benefits. Raskolnikov achieves this higher conviction probability by removing some existing legal protections in resolving taxpayer disputes with the IRS and introducing more stringent rules for tax preparers. Taxpayer benefits in CR would include respectful, friendly treatment and some pre-filing assistance for questionable positions. The DR is characterized by a higher penalty rate, no change in the conviction probability, and no special benefits. While Raskolnikov acknowledges that some gamers will prefer CR to DR, rendering the separation incomplete, he argues that some separation is better than no separation and that experimentation could reveal the policy settings most productive of the desired separation. Note that, as in Falkinger and Walther, Raskolnikov does not use the audit rate as a part of the separation mechanism. In discussing whether to use the audit rate to discriminate between gamers and non-gamer, he first argues that amongst gamers, audits constitute the best deterrent. On account of this assertion, the audit rate should be set higher in DR. Doing so will prevent expensive, wasteful audits of CR taxpayers, and prevent crowding out their voluntary compliance. On the other hand, Raskolnikov argues that the CR is likely to include some gamers along with non-gamers (i.e., the separation will be incomplete). Raising the audit rate in CR will enable identifying gamers and support a more complete separation, pushing the gamers to choose DR instead. Given these two opposing arguments, Raskolnikov asserts that a uniform audit rate could be a credible government policy under his system.

Since taxpayers engage in compliance decisions on a regular basis (e.g., annually), one wonders whether the Falkinger-Walther and Raskolnikov separations would survive repeated choices over time.[2] If the authority exploits the separation to target enforcement (e.g., audits), then it would set a higher audit rate in Falkinger-Walther's NO BONUS and Raskolnikov's deterrence regimes. When making a regime choice in the following year, taxpayers would likely incorporate their prior audit experience. In the second year, a big fish who had been audited in NO BONUS would knowingly face a higher audit rate as well as a lower penalty rate, so that her expected penalty could be smaller or larger than if she moved to the BONUS regime. This might also apply to a big fish who had not been audited, if audit rate information becomes public. However, so long as the authority sets the penalty and audit rate differentials to maintain the separation (high enough penalty rate in BONUS and low enough audit rate differential), the Falkinger-Walther separation could persist (small fish who are aware of the audit rate differential would have an additional reason for choosing the BONUS regime). For the Raskolnikov proposal, auditing gamers in the deterrence regime visits a double whammy on these taxpayers in the second year: a higher likelihood of being audited and a bigger penalty. Thus, it is harder to imagine that the separation could persist, as these gamers would then have an incentive to move to the compliance regime (lower audit rate and lower penalties). Suppose instead that, under the (incomplete) Raskolnikov separation, the authority targets audits on CR taxpayers. In the following year, a CR gamer who was audited but faced a lower penalty rate might rationally decide to stay put. In this case, time reinforces an incomplete separation. However, it is also true that resources are wasted on compliant CR taxpayers.[3]

In a series of dictator game experiments, Lazear, Malmendier and Weber (2012) demonstrate that people who self-select into a sharing regime behave differently from those who participate without an express opportunity to opt out. Initially using a between-subjects design, the authors find that the number of subjects who share a portion of their endowment in a standard dictator game regime (with no opportunity to self-select either into or out of the game) is more than twice as high as the number who share when self-selection is allowed. Then, employing a within-subjects design, they show that some subjects (33–41 percent in their experiments) who share when not allowed to opt out of a dictator game will, given a self-selection opportunity, prefer to opt out and not to share. Applied to the tax compliance situation we study here, these results suggest that the compliance behavior of taxpayers who sort themselves into particular compliance regimes may be quite different from that of an unsorted population of taxpayers.

The Experiment

We conducted 16 one-hour experiment sessions in April 2012 at a computer lab on the University of Minnesota's Minneapolis campus.[4] The lab's 44 networked computer workstations are installed in high-sided carrels that partially isolate subjects from one another. At the beginning of each session, subjects selected carrels, and heard an overview of the experiment and data confidentiality procedures read aloud. Experiment instructions appeared on the computer screens, and subjects had paper copies of the same instructions at their workstations. Subjects also were given paper and pencils to perform calculations.

The experiment proceeded in three treatment stages, with multiple rounds in each stage. All subjects progressed through the stages in the same order. Each stage began with a screen providing the instructions for that treatment, including the number of rounds to be completed. That screen was followed by a three-question, multiple-choice quiz on the instructions. A subject who answered a quiz question incorrectly was informed of the correct answer. The quiz scores in each round were retained as input into the subject's payout. In each treatment, the quiz was followed by at least one practice round.

Within each treatment, subjects faced the possible experiment states—combinations of experiment parameters—in random order. The parameter values for each state are listed in Table 1. At the beginning of each round, the computer screen showed the value of the parameters for that round. Subjects then earned income by completing tasks on the computer terminal within a fixed period of time.[5] The earning task was the same for each round. A list of 30 English words appeared in random order in a box on the computer screen. Subjects completed tasks by correctly sorting words into one of three alphabetical bins: A-H, I-P, and Q-Z. Their earnings for a round equaled the wage rate ($1.20) times the number of correctly sorted words. In all but the third treatment practice rounds, the earning time was 30 seconds. At the conclusion of the earning time, the number of completed tasks appeared on the screen. When each of the three treatments ended, the computer program randomly selected one round (excluding practice rounds) to use to determine the payout for that treatment. Each round in a treatment had the same chance of being selected to be a "payout round." The randomly selected round was not revealed to subjects. At the end of the experiment session, subjects received cash payouts equal to 50 percent of the credited amounts from the three randomly chosen payout rounds plus $0.50 for each correctly answered quiz question, rounded up to the nearest dollar. The average payout was $34.

In each treatment, subjects were charged a tax equal to a tax rate times their income. In the first treatment, subjects reported their earnings to the experiment authority, which randomly audited a fraction of the participants to verify their reported income. Subjects who were audited and found to have reported less income than they earned were charged the correct tax amount: the tax rate times the amount they actually earned. In addition, they were charged a penalty equal to the penalty rate times the tax due on underreported income (penalty rate x tax rate x amount of underreported income). There was no penalty if a report was audited and the subject accurately or overstated income. However, overstated income was corrected (so overstating income did not result in an increased payout for a round), and the tax was recalculated using the correct amount of earned income. If a report was *not* audited and the subject either understated or overstated income, there was no penalty and the tax was based on reported income.

Prior to the earning period in each round of the first treatment, subjects were informed of the tax rate, audit probability, and penalty rate. They reported the earnings by typing a dollar amount into a box. Prior to clicking "submit," the screen showed the subject's tax, penalty, and net-of-tax-and-penalty earnings if audited, and the tax and net-of-tax earnings if not audited for the tentative reported amount of earnings. Subjects could, therefore, experiment with different amounts of reported earnings prior to finalizing their submission.

Subjects proceeded through 12 rounds of the first treatment. Each subject faced tax rates of .15 and .30; audit rates of zero, .10, and .50; and penalty rates of 1.0 and 2.0 in random order.

TABLE 1. Summary of Experiment States

State	Treatment	Tax rate	Probability of audit	Penalty rate	Burden reduction fee	Form type
1	1	0.15	0.00	1.00		
2	1	0.15	0.10	1.00		
3	1	0.15	0.50	1.00		
4	1	0.15	0.00	2.00		
5	1	0.15	0.10	2.00		
6	1	0.15	0.50	2.00		
7	1	0.30	0.00	1.00		
8	1	0.30	0.10	1.00		
9	1	0.30	0.50	1.00		
10	1	0.30	0.00	2.00		
11	1	0.30	0.10	2.00		
12	1	0.30	0.50	2.00		
13	2	0.15			$0.25	Long
14	2	0.15			$0.50	Long
15	2	0.15			$1.00	Long
16	2	0.15			$2.00	Long
17	2	0.15			$4.00	Long
18	3	0.15	0.00	2.00		None
19	3	0.15	0.10	2.00		None
20	3	0.15	0.50	2.00		None
21	3	0.15	0.00	2.00		Short
22	3	0.15	0.10	2.00		Short
23	3	0.15	0.50	2.00		Short
24	3	0.15	0.00	2.00		Long
25	3	0.15	0.10	2.00		Long
26	3	0.15	0.50	2.00		Long

In the second treatment, reporting of earnings required more time and effort. Prior to reporting their earnings, subjects had to calculate their total earnings and enter three data items into an on-screen form. Two of those items (number of tasks completed and seconds per task) appeared on the screen. The third, their workstation number, was posted on their carrel. Therefore, burden occurred in several ways: a calculation, locating information, and entering data.

At the beginning of each round, subjects had the option of paying a fee to be relieved of the reporting burden. If they paid the fee, they didn't have to enter any data or complete any calculations, and their earnings were automatically reported to the authority. Subjects who paid the burden reduction fee were able to move on to the third stage more quickly, ultimately completing the entire experiment in less time.

In this treatment, inaccurate earning reports were not accepted. Instead, subjects who did not pay the burden reduction fee repeated the earnings report until they entered the correct amount. Therefore, there was no opportunity to cheat in the second treatment, reports were not audited, and there were no misreporting penalties.

There were five rounds in this treatment, each with a tax rate of .15. Each subject faced burden reduction fees of $0.25, $0.50, $1.00, and $2.00, and $4.00 in random order.

In the third treatment, subjects chose between two reporting regimes. In the automatic-reporting (transparent, variable burden) regime, actual earnings were automatically reported to the experiment authority, removing any opportunity to underreport. The reporting burden was either none, a short form (requiring the subject to enter only the workstation number), or a long form with the same requirements as in treatment two. The self–reporting (non-transparent, high burden) regime always imposed the highest level of burden. However, subjects *chose* the amount of earnings they reported to the authority, giving them the opportunity to underreport. Reports were randomly audited, and penalties were applied as in the first treatment.

To ensure that subjects understood the differences among the burden levels, the third treatment included three practice rounds (with shortened earning periods to save time), each presenting the self-reporting regime with a different form type.

Prior to making the regime choice at the beginning of each round, subjects were informed of the tax rate, the form type in the automatic-reporting regime, and the audit and penalty rates in the self-reporting regime. There were nine rounds in this treatment, each with a tax rate of .15 and a penalty rate in the self-reporting regime of 2.0. Each subject faced self-reporting regime audit rates of zero, .10, and .50; and automatic-reporting regime form types of none, short, and long in random order.

After completing all rounds in all three treatments, subjects responded to a survey requesting demographic data, their perceptions of the experiment, and their responses to questions that identify their basic personality types. They also answered a single question about their views on tax compliance that is patterned on a question appearing in the World Values Survey (2011): Is cheating on your taxes, if you have a chance, always justifiable, never justifiable, or something in between?[6]

Subject Recruitment and Characteristics

Subject Recruiting

During March and April 2012, we recruited subjects from among students and staff at the University of Minnesota and from the general public in the Minneapolis area. We recruited via flyers on campus, visits to undergraduate classes, ads on Craig's List™ and in the campus newspaper *The Minnesota Daily*, and emails to University students and staff.

As subjects received their payouts and exited the computer lab, we asked them how they learned about the experiment. The most common way of learning about the experiment was via word of mouth (33 percent of subjects) with that share varying from a low of zero percent of subjects in the third and fourth sessions to a high of 59 percent of the participants in session 15. Fairly equal shares of all subjects (24 and 21 percent, respectively) learned about the experiment from a flyer posted on campus or Craig's List ™. Twelve percent of subjects responded to an email to University students and staff, and 7 percent were recruited during a visit to their class.

Subject Characteristics and Perceptions

The result of our recruitment strategy was a subject pool that is more diverse than is typical for on-campus experiments. Subjects' demographic characteristics are summarized in Table 2. Slightly more than half (52 percent) of our subjects were female, 38 percent were non-White, 24 percent were age 30 or older, and 28 percent were not students. Of the student subjects, 68 percent had majors outside of economics or related fields. Students were roughly evenly distributed among years in school, from freshman to graduate.

To see how subjects' behavior in the experiment might correlate with their work circumstances, and to compare our subject pool to the general population, we asked about employment, occupation, and status as a business owner or partner. Their responses are also summarized in Table 2. Nearly 64 percent of our subjects reported being employed, which is a larger share than that for the U.S. over-age-16 population (58.4 percent in April 2012).[7] About 32 percent of the employed subjects chose "other" for their occupation. Of the remaining employed subjects, 20 percent each chose management and professional, service, or sales and office. In comparison, a larger share of the general employed U.S. population is in management and professional occupations (38 percent in 2011), and smaller shares are in service (18 percent) and sales and office occupations (24 percent). Among both our subjects and the general U.S. population, less than 10 percent are employed in occupations related to natural resources/construction/maintenance or production/transportation/material moving.[8]

TABLE 2. Subject Characteristics with Mean Index Value

Subject characteristics	Number of observations	Percent of total	Propensity to cheat		Willingness to pay for burden reduction
			Index 1	Index 2	
Female	173	52.4	.46	.28	.15
Male	157	47.6	.48	.32	.18
Age: under 20	61	18.5	.61	.39	.14
Twenties	191	57.9	.47	.32	.14
Thirties	33	10.0	.45	.28	.17
Forties	14	4.2	.30	.15	.24
Fifties	21	6.4	.21	.09	.27
Sixties	10	3.0	.32	.11	.40
Race/ethnicity: White	205	62.1	.46	.30	.17
Asian/Pacific Islander	76	23.0	.56	.35	.13
Black/African American	25	7.6	.40	.19	.16
Hispanic/Latino	9	2.7	.28	.18	.16
American Indian	3	1.0	.58	.25	.60
Other	12	3.6	.37	.23	.08
Student	239	72.4	.50	.32	.15
Not a student	91	27.6	.38	.23	.20
Year in school[1]: Freshman	51	21.3	.60	.37	.15
Sophomore	48	20.1	.51	.32	.16
Junior	42	17.6	.51	.34	.13
Senior	51	21.3	.43	.27	.19
Graduate student	47	19.7	.46	.32	.11
Major[1]: Not economics or related	151	63.2	.50	.31	.16
Economics or related	77	32.2	.48	.34	.13
Undeclared major	11	4.6	.68	.38	.11
Employed	211	63.9	.44	.28	.16
Unemployed	119	36.1	.52	.33	.16
Occupation[2]: Management/professional	43	20.4	.39	.23	.15
Service	43	20.4	.47	.32	.13
Sales and office	42	19.9	.47	.28	.21
Natural resources/construction/maintenances	10	4.7	.54	.34	.42
Production/transportation/material moving	6	2.8	.42	.28	.09
Other	67	31.8	.42	.29	.14
Business owner or partner	34	10.3	.47	.30	.16
Not a business owner or partner	296	89.7	.43	.26	.18

[1] Year in school and major are for students only.
[2] Ocupation is for employed subjects only.

The experiment subjects' perceptions of the experiment are summarized in Table 3. A majority of subjects reported that they agree or strongly agree that they understood the experiment instructions (81 percent) and that their anonymity would be preserved (84 percent). Both questions had fewer than six percent of "disagree" and "strongly disagree" responses.

To compare our subjects' behavior in the experiment to their self-reported views of tax compliance, we asked them to choose from a 10-point Likert scale whether cheating on their taxes, if given the chance, is always justifiable (1), never justifiable (10), or somewhere in between. Nearly 29 percent of subjects chose never justifiable, and about 5 percent chose always. (In response to a similar question in the 2006 World Values Survey, 60 percent of the 1,249 U.S. respondents chose never justifiable, and 1 percent chose always.)[9]

TABLE 3. Perception Statement Responses

(Statements in Italics)

Response	Number of observations	Percent of total
I understood the instructions for this experiment.		
Strongly Agree (5)	134	40.6
Agree (4)	133	40.3
Neutral (3)	44	13.3
Disagree (2)	6	1.8
Strongly Disagree (1)	13	3.9
The procedures used in this study will preserve my anonymity.		
Strongly Agree (5)	188	57.0
Agree (4)	89	27.0
Neutral (3)	39	11.8
Disagree (2)	3	1.0
Strongly Disagree (1)	11	3.3
Is cheating on your taxes, if you have the chance, always justifiable, never justifiable, or something in between?		
Never justifiable (10)	95	28.8
(9)	43	13.0
(8)	42	12.7
(7)	30	9.1
(6)	16	4.9
(5)	43	13.0
(4)	25	7.6
(3)	14	4.2
(2)	7	2.1
Always justifiable (1)	15	4.6

Results

Determinants of Underreporting

The first treatment included 12 rounds in which subjects earned income and reported their earnings to the experiment authority. Reports were randomly audited, and subjects who were audited and found to have underreported income were charged the correct amount of tax and penalized. The experiment design exhibited within-subject variation in the tax, audit and penalty rates and between-subject variation in demographic characteristics, propensity to cheat and, in the second and third treatments, willingness to pay for burden reduction. The design allows us to estimate subjects' propensity to underreport income and to pay for burden reduction, and to identify the determinants of the amount of underreported income. Table 4 shows the share of subjects who were noncompliant and the mean amount of underreported income for the noncompliant subjects by experiment state.

TABLE 4. Compliance, Burden Reduction, and Regime Choice by Experiment State

State	Percent of subjects noncompliant[1]	Mean underreported amount[2]	Percent of subjects choosing burden reduction	Percent of subjects choosing self-reporting
1	62.7	$18.42		
2	53.3	$13.91		
3	27.0	$8.60		
4	63.3	$18.15		
5	48.5	$12.98		
6	23.6	$7.24		
7	63.9	$18.75		
8	54.2	$13.71		
9	27.3	$7.45		
10	65.8	$18.04		
11	48.8	$12.32		
12	23.6	$6.75		
13			42.1	
14			32.1	
15			25.2	
16			16.4	
17			10.3	
18	78.3	$24.44		72.7
19	70.2	$18.28		57.9
20	32.8	$4.79		35.2
21	80.7	$24.16		72.4
22	69.0	$18.07		59.7
23	27.3	$7.35		36.7
24	82.9	$23.90		70.9
25	67.0	$18.13		60.6
26	28.0	$8.10		35.8

[1] For states 18-26, the percent noncompliant is among subjects who chose the self-reporting regime.

[2] The mean underreported amount is for noncompliant subjects only.

Table 5 shows the results of two Tobit[10] regressions with the amount of underreported income as the dependent variable. Results in the first column are from a fixed effects model (using individual subject dummy variables as controls), and results in the second column are from a model using subjects' responses to our demographic questionnaire as controls. In both models, the tax rate had a positive but insignificant impact on the level of underreporting while both the audit rate and the penalty rate had significantly negative impacts. Because income was determined by subjects' performance in the earning task, it varied both across subjects and across rounds for any individual subject. In both models, underreporting varied positively with income, with another $1 of earned income being associated with additional underreporting of $0.20 or $0.37, holding all else equal (for fixed-effects and demographic controls, respectively).

In the second model, several of the demographic control variables were significantly correlated with the amount of underreported income. Males underreported more than females. Relative to Whites, Blacks and Other Races underreported less. Relative to subjects under age 20, all other age groups underreported less.

TABLE 5. Coefficients from Tobit Regression of Treatment 1 Results

Dependent Variable: The amount of underreported income

Control variables	Fixed effects	With demographic controls
Tax rate	2.525 (3.853)	1.710 (5.248)
Audit rate	-54.418 (1.657)***	-54.055 (2.189)***
Penalty rate	-2.129 (0.577)***	-2.067 (0.787)**
Income	0.652 (0.120)***	1.234 (0.107)***
Female		-3.531 (0.834)***
Age 60+		-12.208 (3.410)***
Age 50s		-16.589 (3.001)***
Age 40s		-12.200 (3.105)***
Age 30s		-5.487 (2.451)***
Age 20s		-6.994 (2.035)***
Black		-5.245 (1.637)***
Asian		0.996 (1.156)
Other races		-6.927 (1.687)***
Sophomore		3.939 (2.062)
Junior		6.161 (2.495)
Senior		-0.435 (2.451)
Graduate student.		3.794 (2.471)
Major Economics or related		-1.291 (1.071)
Major undeclared		5.058 (2.254)**
Non-student		2.208 (2.511)

Footnotes at end of table.

TABLE 5. Coefficients from Tobit Regression of Treatment 1 Results—Continued

Dependent Variable: The amount of underreported income

Control variables	Fixed effects	With demographic controls
Business owner		1.907 (1.440)
Management/professional		0.166 (1.336)
Natural res./constr./maint.		-1.777 1.658
Other occupation		5.004 (2.510)**
Production/transport.		-2.353 (1.471)
Sales&office		-0.998 (3.344)
Service		0.014 (1.596)
Round	0.378 (0.097)***	0.117 (0.125)
WTP index		-2.154 (1.548)
Cheating not justifiable[1]		-1.317 (0.161)***
Constant	-78.753 (179.412)	-0.706 (3.479)
Number of observations	3,950	3,950

* p<0.1; ** p<0.05; *** p<0.01 (standard errors in parentheses)

[1] "Cheating not justifiable" is the subject's response to the question "Is cheating on your taxes, if you have the chance, always justifiable, never justifiable, or something in between?" where 1=never justifiable and 0=always justifiable.

Note: The marginal impact on observed underreporting is the Tobit regression estimate x the probability of underreporting a fraction of the income. In Treatment 1 the average probability was .302.

Subjects' response to the question, "Is cheating on your taxes, if you have the chance, always justifiable, never justifiable, or something in between?" was significantly correlated with underreporting. A higher numerical response—meaning a *lower* tendency to justify cheating—was associated with lower levels of underreported income.

Repeating these Tobit regressions and substituting the proportion of income underreported in place of the level of underreporting, the results remain qualitatively the same (see Table 5a). Since these regressions also included income as an explanatory variable, it is possible to explore its impact on underreporting quadratically. Notice that in both models the income coefficient is both positive and significant, suggesting that the marginal impact on compliance of earning another $1 increases as a subject earns more.

Measures of Propensity to Cheat

Observations of subject behavior over the 12 rounds of the first treatment provide several ways to measure propensity to underreport income. In this paper, we derive two simple propensity-to-cheat indices.[11] The first index (PTC1) is the share of rounds (out of 12) a subject reports income that is strictly less than their actual earnings. The index score varies from zero to one, with a subject who never underreports getting a score of zero, and a subject who underreports at every opportunity a score of 1.0. The mean value of PTC1 over all subjects was .47, with 23 percent of subjects receiving scores of zero, and 10 percent of subjects receiving scores of 1.0.

TABLE 5a. Coefficients from Tobit Regression of Treatment 1 Results

Dependent Variable: the proportion of underreported income

Control variables	Fixed effects	With demographic controls
Tax Rate	0.125 (0.167)	0.076 (0.230)
Audit Rate	-2.344 (0.072)***	-2.350 (0.096)***
Penalty rate	-0.089 (0.025)***	-0.087 (0.0345)**
Income	0.011 (0.005)**	0.038 (0.005)***
Female		-0.153 (0.037)***
Age 60+		-0.584 (0.149)***
Age 50s		-0.767 (0.131)***
Age 40s		-0.567 (0.135)***
Age 30s		-0.245 (0.107)**
Age 20s		-0.319 (0.089)***
Black		-0.241 (0.072)***
Asian		0.043 (0.051)
Other races		-0.319 (0.074)***
Sophomore		0.171 (0.090)*
Junior		0.284 (0.109)***
Senior		-0.020 (0.107)
Grad.		0.168 (0.108)
Major Economics or related		-0.064 (0.047)
Major undeclared		0.233 (0.099)**

Footnotes at end of table.

TABLE 5a. Coefficients from Tobit Regression of Treatment 1 Results—Continued

Dependent Variable: the proportion of underreported income

Control variables	Fixed effects	With demographic controls
Non-student		0.100 (0.110)
Business owner		0.077 (0.063)
Management/professional		0.015 (0.059)
Natural resources/construction/ maintenance		-0.075 (0.073)
Other occupation		0.241 (0.110)**
Production/transportation		-0.112 (0.065)*
Sales&office		-0.060 (0.146)
Service		0.004 (0.070)
Round	0.017 (0.004)***	0.005 (0.005)
WTP index		-0.095 (0.0676)
Cheating not justifiable[1]		-0.058 (0.007)***
Constant	-3.137 (12.666)	0.3783 (0.152)**
Number of observations	3,950	3,950

* p<0.1; ** p<0.05; *** p<0.01 (standard errors in parentheses)

[1] "Cheating not justifiable" is the subject's response to the question "Is cheating on your taxes, if you have the chance, always justifiable, never justifiable, or something in between?" where 1=never justifiable and 0=always justifiable.

Note: The marginal impact on observed underreporting is the Tobit regression estimate x the probability of underreporting a fraction of the income. In Treatment 1 the average probability was .302.

The second index (PTC2) is the share of total income over all 12 rounds that a subject failed to report: [(total actual income—total reported income)/total actual income]. A subject who never underreported received a PTC2 score of zero, and a subject who maximized underreporting by never reporting any income would earn a 1.0. Over all subjects, the mean value of PTC2 was .30. As with PTC1, 23 percent of subjects received scores of zero, but no subjects received PTC2 scores of 1.0. The maximum value of PTC2 was .84. Table 2 (presented earlier) contains separate mean index scores for the subjects, by demographic group.

Table 6 shows the correlation coefficient between the propensity-to-cheat indices and an index constructed from subjects' responses to the justifiability of cheating question, where a value of one corresponds to always justifiable and value of zero corresponds to never justifiable. Correlations range from .22 to .26 depending on the type of correlation coefficient used. Although these are positive and significantly different from zero, these correlations are quite small and suggest that most of the variation in within-experiment propensity to cheat is coming from sources other than subjects' reported feelings about tax cheating.

TABLE 6. Correlations Between Justifiability of Cheating and Propensity To Cheat Indices

All subjects						
	Pearson[1]			Spearman[2]		
	Cheating justifiable	PTC1	PTC2	Cheating justifiable	PTC1	PTC2
Cheating justifiable	1.00			1.00		
PTC1	0.24	1.00		0.26	1.00	
PTC2	0.22	0.78	1.00	0.25	0.79	1.00
Excluding subjects who always complied						
	Pearson			Spearman		
	Cheating justifiable	PTC1	PTC2	Cheating justifiable	PTC1	PTC2
Cheating justifiable	1.00			1.00		
PTC1	0.15	1.00		0.13	1.00	
PTC2	0.13	0.58	1.00	0.12	0.55	1.00

[1]The Pearson product-moment correlation detects a linear relationship between two variables.
[2]The Spearman correlation coefficient detects a monotonic relationship between two variables.

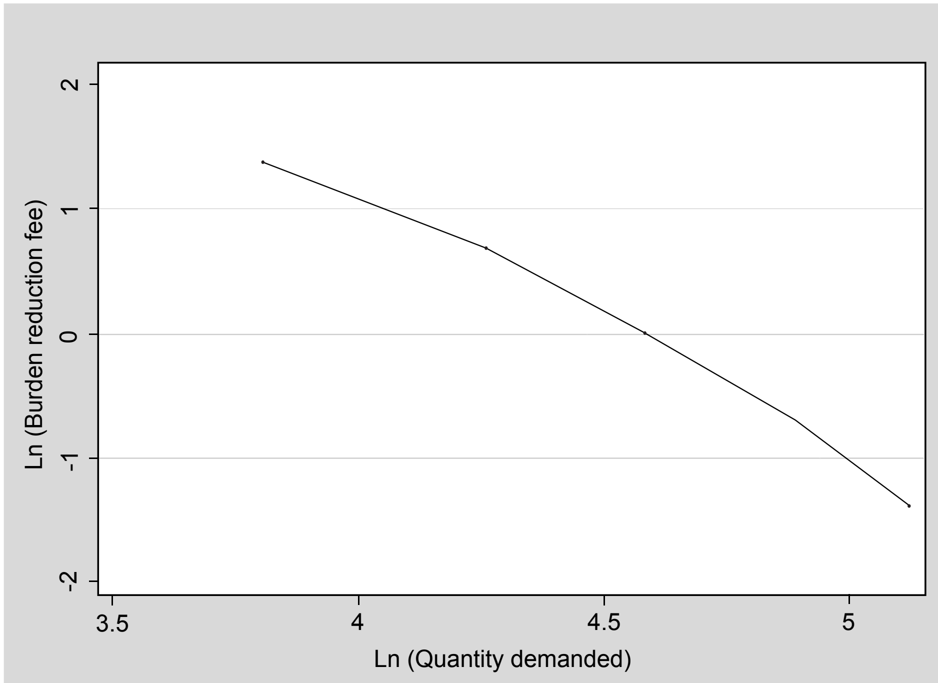
Measures of Willingness To Pay for Burden Reduction

The five rounds of the second treatment were designed to measure the subjects' willingness to pay a fee in order to reduce the burden of reporting their earnings to the experiment authority. That burden had several time and task components. First, subjects had to complete an on-screen Earnings Calculation Worksheet, entering several pieces of information: their laboratory carrel number (posted on the carrel wall), the average time spent completing a task and the number of tasks completed (available in a simultaneously displayed on-screen box), and their calculation of total earnings ($1.20 times the number of completed tasks). Subjects had to continue working on this until all of their entries were correct. Second, an Earnings Reporting Form appeared on the screen. Here, subjects had to enter their total earnings again. If a subject made an incorrect entry, she was asked to try again. Upon a correct total earnings entry, the computer screen displayed "Report Accepted" along with net-of-tax earnings for the round. Alternatively, subjects who paid the burden reduction fee were relieved of completing both of these forms, as the computer screen displayed "Report Accepted" and their earnings, net of tax and the fee, as the round concluded. A different burden reduction fee ($4, $2, $1, $0.50, $0.25), in random order, applied to each round. Table 4 (presented earlier) shows the share of subjects who chose to purchase the burden reduction fee by experiment state.

A subject who chose to pay the burden reduction fee in all five rounds would reduce their earnings in this stage by $7.75, while a subject who did not would have no such earnings reduction. As an index of willingness to pay for burden reduction, we calculated how much each subject paid, as a proportion of $7.75. Index scores therefore run from 0 (no fees paid) to 1.0 (all 5 fees paid). Over all subjects, the mean index score was .16, with 48 percent of subjects scoring zero and 6 percent of subjects scoring 1.0. Table 2 (presented earlier) contains separate mean index scores for the subjects, by demographic group.

As subjects face different fees, their decisions regarding whether to purchase burden reduction trace out a willingness to pay or demand for burden reduction schedule. Our results demonstrate that demand is downward-sloping; more subjects purchased burden reduction when the fee was low than when it was high. Moreover, we find that demand is more price-sensitive (elastic) at higher fees (Figure 1 graphs the schedule with a double logarithmic scale).

FIGURE 1. Willingness To Pay (Demand) for Burden Reduction



Determinants of Regime Choice

In each round of the third experimental treatment, subjects face the same tax rate and penalty rates (0.15 and 2, respectively), but differing audit rates (0, 0.10, and 0.50) and self-reporting forms (no form, short form, long form). Given these round parameters, subjects choose between two reporting regimes: one in which they self-report their earnings and another in which earnings are automatically reported by the computer. They then engage in the same income-producing activity. Subjects choosing the self-reporting regime use the indicated form to report their earnings, and face a possible audit and penalty. Those choosing the automatic regime do not report their earnings and face no audit or penalty. Table 4 shows the share of subjects choosing the self-reporting regime by experiment state.

Table 7 contains the results of a series of probit regressions in which the dependent variable is "1" if the subject chose the self-reporting regime and "0" if she chose automatic reporting. These regressions explore the impact of the audit rate, the form type, subjects' willingness to pay for burden reduction (from Treatment 2), subjects' propensity to cheat (from Treatment 1), and subject demographic characteristics on their regime choice. We display here the calculated marginal effects of the independent variables, with their standard errors in parentheses.

Column 1 measures propensity to cheat as the proportion of Treatment 1 rounds in which a subject's reported income was strictly smaller than actual income (PTC1). Column 3 uses instead PTC2, the proportion of the subject's Treatment 1 income that she underreported. Columns 2 (PTC1) and 4 (PTC2) add the demographic characteristics. The marginal effects we discuss below are all subject to holding everything else constant.

Regarding the audit rate, in all four regressions, subjects were significantly less likely to choose the self-reporting regime as the audit rate increased, confirming our expectations. The magnitudes of the marginal effect are about the same: a 0.1 (10 percentage point) increase in the audit rate is associated with a .06 (6 percentage point) decline in the probability of choosing self-reporting.

TABLE 7. Marginal Effects from Probit Regression for Choice of Self-Reporting Regime in Treatment 3

	(1)	(2)	(3)	(4)
AuditRate	-0.646 (0.032)***	-0.648 (0.032)***	-0.645 (0.032)***	-0.645 (0.031)***
SimpleForm	0.010 (0.020)	0.010 (0.020)	0.011 (0.020)	0.011 (0.020)
LongForm	0.002 (0.020)	0.002 (0.020)	0.002 (0.020)	0.003 (0.020)
Income	0.015 (0.002)***	0.012 (0.002)***	0.014 (0.002)***	0.012 (0.002)***
WTP	-0.264 (0.031)***	-0.257 (0.030)***	-0.273 (0.031)***	-0.268 (0.031)***
PTC1	0.235 (0.023)***		0.221 (0.024)***	
PTC2		0.402 (0.032)***		0.392 (0.034)***
Number of observations	2,970	2,970	2,970	2,970

* p<0.1; ** p<0.05; *** p<0.01 (standard errors in parentheses)

Note: Column 1 includes PTC1 as a regressor and no demographics; Column 2 includes PTC2 and no demographics; Column 3 includes PTC1 and all demographic controls; Column 4 includes PTC2 and all demographic controls

Subjects with higher propensities to cheat were significantly more likely to choose the self-reporting regime, as expected. For the PTC1 index, a 0.1 (10-percentage-point) increase in the proportion of Treatment 1 rounds in which a subject underreported was associated with a 0.02 (2-percentage-point) increase in the probability of choosing self-reporting. The effect, also significant, was about twice as large for PTC2: a 0.1 (10 -percentage-point) increase in the proportion of Treatment 1 income underreported was associated with a 0.04 (4-percentage-point) increase in the probability of self-reporting.

The impact of subjects' willingness to pay for burden reduction was also as expected: those with higher willingness to pay for burden reduction index values were significantly less likely to choose self-reporting. A 0.1 (10-percentage-point) increase in the index is here associated with about a .026 (2.6-percentage-point) decrease in self-reporting.

Subjects who earned more income were significantly more likely to choose self-reporting: a $1 increase in income was associated with a .01 (1-percentage-point) increase in the probability of self-reporting.

Regarding the effect of the type of self-reporting form (none, short, or long), subjects who faced a short form were more likely to choose self-reporting than those who faced a long form (both relative to those facing no form), but neither effect was statistically significant.

In columns 3 and 4, we note that the inclusion of demographic variables has little impact on the coefficients discussed above. Several of these do appear to be statistically significant correlates of self-reporting. Males were more likely than females to choose self-reporting. Relative to Whites, Blacks (column 4 only) and Other Races were more likely to self-report while Asians were less likely to do so (column 3 only). Only one of the age categories had a significant impact: subjects 60 years of age and older were less likely to self- report, relative to those 20-30 (column 3 only). Year in school, not having declared a major and not being a student had no significant impacts on regime choice. Subjects who were majors in Economics or a related field were significantly less likely to choose the self-reporting regime (column 4 only), while those who owned businesses were significantly more likely to self-report than non-business owners. Regarding the occupational dummies, Management & Professional, Sales & Office, Service and Other all had significantly negative coefficients. This suggests that subjects in these occupations were less likely to select the self-reporting regime, relative to those who were not employed.[12]

Determinants of Underreporting in the Self-Reporting Regime

Table 8 contains the results of a series of linear regressions that explore the behavior of those subjects who chose the self-reporting regime in Treatment 3. In all four regressions, the dependent variable is a subject's amount of underreported income. The first two columns do not include demographic characteristics, while the last two columns do (demographic variables suppressed here). Columns 1 and 3 use the PTC1 measure of cheating propensity; columns 2 and 4 use the PTC2 alternative measure.

Both propensity-to-cheat indices are significantly associated with greater underreporting. A 0.1 increase in PTC1 (proportion of Treatment 1 rounds with underreporting) is related to $1.24 more Treatment 3 underreporting. The same increase in PTC2 (proportion of stage 1 income underreported) has a much larger impact ($2.40). Both of these results persist in the regressions that include demographic characteristics. Together with the findings of the prior section, these results suggest that subjects who have a greater propensity to cheat are more likely to choose a self-reporting regime and, thereafter, to underreport more of their income.

TABLE 8. Coefficients from a Linear Regression in Treatment 3

Dependent Variable: Amount Underreported, conditional on choosing the self-report regime

	(1)	(2)	(3)	(4)
Audit rate = 0.1	-7.628 (0.537)***	-7.501 (0.484)***	-7.472 (0.526)***	-7.399 (0.478)***
Audit rate = 0.5	-16.911 (0.628)***	-15.933 (0.567)***	-16.317 (0.618)***	-15.544 (0.562)***
Income	0.785 (0.050)***	0.601 (0.046)***	0.634 (0.057)***	0.525 (0.052)***
WTP	5.993 (1.121)***	6.087 (1.009)***	5.440 (1.146)***	5.712 (1.041)***
PTC1	12.391 (0.738)***		12.125 (0.766)***	
PTC2		23.954 (0.883)***		23.362 (0.917)***
Constant	-8.942 (1.415)***	-6.145 (1.257)***	-4.518 (2.166)**	-2.432 (1.963)
R2	0.46	0.56	0.49	0.58
Number of observations	1,656	1,656	1,656	1,656

* p<0.1; ** p<0.05; *** p<0.01 (standard errors in parentheses)

Note: Column 1 includes PTC1 as a regressor and no demographics; Column 2 includes PTC2 and no demographics; Column 3 includes PTC1 and all demographic controls; Column 4 includes PTC2 and all demographic controls

The willingness to pay for burden reduction index also has a positive and significant coefficient. A 0.1 increase in the index (from Treatment 2 behavior) is associated with approximately another $0.60 of underreporting in Treatment 3, both with and without the demographic variables. As we note in the previous section, subjects with a higher willingness to pay for burden reduction were less likely to choose the self-reporting regime. However, the results here imply that, among the subjects who *chose* self-reporting, a higher willingness to pay for burden reduction—or a greater distaste for burden—leads to larger amounts of underreporting. This can be interpreted as meaning that those willing to pay more for burden reduction *that we see choosing self-reporting* are even more willing to trade burden for the additional take-home pay that results from cheating. Further work would be required to determine if this result persists in a model that controls for selection bias (regime choice) and/or fixed effects.

In these regressions we entered the audit rate as a categorical variable (rates of 0.1 and 0.5), omitting the zero audit rate category. As we would expect, subjects underreported less when the audit rate rose to either 0.10 or 0.50. The 10-percent audit rate was associated with between $7.40 and $7.63 less underreporting, relative to a zero audit rate; that comparison range was $15.54–$16.91 for a 50-percent audit rate.

A self-reporting subject's actual income in a Treatment 3 round also appeared to affect the magnitude of her underreporting. A $1 increase in income is associated with between $0.53 and $0.79 in additional underreporting, holding all else equal.

Several of the demographic variables have statistically significant effects. Being male has a negative impact on underreporting (column 4) as do being Black, or being Asian (column 3) or having an "other" race (column 3), measured relative to being White. The two oldest age categories are associated with less underreporting (column 3), relative to subjects in their 20's. Regarding student status, sophomores underreported less, relative to freshmen (column 4), while Economics majors underreported more, relative to students declaring other majors. In the previous section, subjects who own their own businesses were more likely to choose the self-reporting regime. However, here those business owners who did self report had significantly lower underreporting than non-business owners. Subjects who work in Natural Resources, Construction and Maintenance or Services or "Other" are associated with significantly higher overreporting, relative to subjects who are not employed.[13]

Table 9 contains the results of Tobit[14] regressions to explain amounts of underreported income, conditional on subjects choosing the self-reporting regime. The first two columns do not include demographic characteristics, while the last two columns do (demographic variables suppressed here). Columns 1 and 3 use the PTC1 measure of cheating propensity; columns 2 and 4 use the PTC2 alternative measure. The results are similar to those reported above for the linear regressions. Both propensity-to-cheat indices are significantly associated with greater underreporting. A 0.1 increase in PTC1 (proportion of Treatment 1 rounds with underreporting) is related to $1.43 more Treatment 3 underreporting. The same increase in PTC2 (proportion of stage 1 income underreported) has a much larger impact ($2.17). The willingness to pay for burden reduction index also has a positive and significant coefficient. A 0.1 increase in the index (from Treatment 2 behavior) is associated with approximately another $0.54 of underreporting without and $0.45 with demographic control variables in Treatment 3.

In these regressions we entered the audit rate as a categorical variable (rates of 0.1 and 0.5), omitting the zero audit rate category. As we would expect, subjects underreported less when the audit rate rose to either 0.10 or 0.50. The 10-percent audit rate was associated with between $5.66 and $6.08 less underreporting, relative to a zero audit rate; that comparison range was $12.71–$14.84 for a 50-percent audit rate.

A self-reporting subject's actual income in a Treatment 3 round also appeared to affect the magnitude of her underreporting. A $1 increase in income is associated with between $0.17 and $0.45 in additional underreporting, holding all else equal.

Finally, we note positive, significant coefficients for the Round variable, with underreporting increasing between $0.23 and $0.30 with each Treatment 3 round.

Table 9a repeats the two-limit Tobit analysis, substituting the proportion of income underreported for the level of underreporting, as the dependent variable. Qualitatively, the results are the same. As in Table 5a, the significantly positive coefficient on income (models 1, 2 and 3) suggests that its marginal impact increases as a subject earns more.

Using Regime Choice To Separate Compliant and Noncompliant Subjects

Above we established that subjects with higher propensities to cheat were significantly more likely to choose the self-reporting regime and having made that choice, underreported significantly more income. Here, we classify subjects according to their Treatment 1 behavior and observe how different classes of subjects behave when faced with a regime choice. We create two classes: subjects who fully reported their income in every round of Treatment 1 (PTC1 = 0) and subjects who at least once underreported their Treatment 1 income (PTC1 > 0).

TABLE 9. Coefficients from Tobit Regression in Treatment 3

Dependent Variable: Amount underreported, conditional on choosing the self-report regime

	(1)	(2)	(3)	(4)
Audit rate = 0.1	-19.692 (1.609)***	-18.514 (3.770)***	-19.285 (1.557)***	-18.327 (1.372)***
Audit rate = 0.5	-48.031 (2.341)***	-42.053 (1.950)***	-45.989 (2.257)***	-41.140 (1.920)***
Income	1.457 (0.150)***	0.866 (0.131)***	0.883 (0.172)***	0.559 (0.152)***
WTP	17.401 (3.302)***	16.665 (2.837)***	14.561 (3.280)***	14.796 (2.864)***
PTC1	46.282 (2.673)***		44.457 (2.661)***	
PTC2		70.292 (3.193)***		67.809 (3.196)***
Round	0.871 (0.273)***	0.738 (0.235)***	0.964 (0.264)***	0.803 (0.230)***
Constant	-42.067 (4.567)***	-26.142 (1.870)***	-27.221 (5.779)***	-18.139 (5.001)***
Scale	23.311	19.804	22.177	19.094
Log Likelihood	-3016	-2880	-2955	-2834
Number of observations	1,655	1,655	1,655	1,655

* *p*<0.1; ** *p*<0.05; *** *p*<0.01 (standard errors in parentheses)

Notes: Column 1 includes PTC1 as a regressor and no demographics; Column 2 includes PTC2 and no demographics; Column 3 includes PTC1 and all demographic controls; Column 4 includes PTC2 and all demographic controls. The marginal impact on observed underreporting is the Tobit regression estimate x the probability of underreporting a fraction of the income. In Treatment 3 the average probability was .309.

With 330 subjects and 9 rounds in Treatment 3, there were 2,970 total opportunities for subjects to make a regime choice. Subjects chose self-reporting 55.8 percent of the time, and those who made that choice underreported 65.4 percent of the time, underreporting an average of $20.24. The share of subjects choosing self-reporting decreased with the self-reporting regime audit rate, from 72 percent when the audit rate was zero to 35.9 percent with a .50 audit rate. And within the self-reporting regime, both the share of subjects underreporting and the average amount underreported decreased with the audit rate.

Over all audit rates, subjects who sometimes underreported in Treatment 1 chose self-reporting 61.2 percent of the time, and those who made that choice underreported 76 percent of the time, with a mean amount of underreporting of $20.45. Over all audit rates, 37.4 percent of subjects who always complied in Treatment 1 chose self-reporting, but few of those (7.4 percent) chose to underreport. The average amount underreported by this group was $8.81.

Subjects who *always* underreported in Treatment 1 (PTC1 = 1) chose the self-reporting regime only 54.9 percent of the time. However, having made that choice, they underreported 92.3 percent of the time. Nevertheless, 45.1 percent of subjects who consistently underreported in Treatment 1 self-selected into a regime that automatically increased their compliance rate to 100 percent.

The number of subjects who chose self-reporting and subsequently underreported income divided by the total number of subjects provides an overall noncompliance rate for Treatment 3 rounds. As shown in Table 10, for rounds in which the audit rate was zero, 10 percent, and 50 percent, the Treatment 3 noncompliance rates were 59 percent, 41 percent, and 10 percent, respectively. The comparable noncompliance rates for Treatment

TABLE 9a. Coefficients from Tobit Regression in Treatment 3

Dependent variable: Proportion of Income underreported, conditional on choosing the self-report regime

	(1)	(2)	(3)	(4)
Audit rate = 0.1	-0.714 (0.058)***	-0.666 (0.050)***	-0.696 (0.056)***	-0.656 (0.049)***
Audit rate = 0.5	-1.729 (0.085)***	-1.499 (0.069)***	-1.646 (0.081)***	-1.459 (0.068)***
Income	0.037 (0.005)***	0.016 (0.005)***	0.015 (0.006)**	0.004 (0.005)
WTP	0.620 (0.119)***	0.585 (0.100)***	0.505 (0.117)***	0.509 (0.101)***
PTC1	1.684 (0.097)***		1.602 (0.096)***	
PTC2		2.557 (0.114)***		2.449 (0.113)***
Round	0.032 (0.010)***	0.027 (0.008)***	0.035 (0.009)***	0.029 (0.008)***
Constant	-1.099 (0.164)***	-0.540 (0.133)***	-0.514 (0.207)**	-0.224 (0.176)
Scale	0.844	0.705	0.798	0.676
Log Likelihood	-1318	-1173	-1253	-1124
Number of observations	1,655	1,655	1,655	1,655

* *p*<0.1; ** *p*<0.05; *** *p*<0.01 (standard errors in parentheses)

Notes: Column 1 includes PTC1 as a regressor and no demographics; Column 2 includes PTC2 and no demographics; Column 3 includes PTC1 and all demographic controls; Column 4 includes PTC2 and all demographic controls. The marginal impact on observed underreporting is the Tobit regression estimate x the probability of underreporting a fraction of the income. In Treatment 3 the average probability was .309.

TABLE 10. Percent of All Subjects Noncompliant and Shares of Unreported Income by Regime and Audit Rate

Audit Rate	Single Regime	Dual Regime
	Percent Noncompliant	
0	63%	59%
10%	49%	41%
50%	24%	10%
	Share of Income Unreported	
0	55%	55%
10%	31%	28%
50%	7%	3%

1 were 63 percent, 49 percent, and 24 percent, respectively. The results are similar for the shares of income unreported. At each audit rate, therefore, both the percentage of noncompliant subjects and the share of income unreported were lower in the dual-regime case than under the single regime.

Conclusion

This paper presents the results of a laboratory experiment designed to explore taxpayers' willingness to trade non-transparency—or the opportunity to underreport income—for a reduction in reporting burden. We measured subjects' propensity for noncompliance by observing their reporting behavior in 12 experimental states that exhibited varying tax, audit, and penalty rates. We also measured subjects' willingness to pay for burden reduction by introducing a reporting burden and offering subjects the option of paying a fee to avoid it. We then observed whether these measures could explain subjects' choice of regime and, for subjects choosing a self-reporting regime, their compliance behavior.

We found that subjects were significantly less likely to choose the self-reporting regime as the audit rate increased. Subjects with higher propensities to cheat and who earned more income were significantly more likely to choose the self-reporting regime. Those with higher willingness to pay for burden reduction were significantly less likely to choose self-reporting.

Conditional on having chosen the self-reporting regime, we found that both propensity to cheat and willingness to pay for burden reduction were significantly associated with greater underreporting.

We classified subjects according to their behavior in a basic tax compliance environment. We found that a majority of subjects who were classified as sometimes noncompliant self-selected into the automatic-reporting regime, eliminating any opportunity for them to underreport. We also found that more than a third of subjects who were classified as fully compliant opted into the self-reporting regime, but once there, rarely chose to underreport. Comparing the share of subjects who were noncompliant under the single regime and the dual-regime system, we found that the noncompliance rate was lower when subjects were offered a choice of regimes.

Acknowledgments

This research was conducted under contract with the Internal Revenue Service, IRS/TIRNO-11-P-00718. The authors thank the IRS for supporting the research, Matthew Bombyk for graduate research assistance, Ryan Anderson for computer programming, and Alan Plumley, Elaine Maag, and Alex Möhlmann for helpful comments.

References

Allingham, M. G., and Sandmo, A. (1972): "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics* 1: 323–338.

Alm, James, Mark B. Cronshaw, and Michael McKee. "Tax Compliance with Endogenous Audit Selection Rules," *Kyklos* 46: 27–45, 1993 Fasc 1.

Alm, James; Michael Jones; and Todd Cherry. "Taxpayer Information Assistance Services and Tax Compliance," Tulane University Working Paper 1101, April 2011.

Andreoni, James; Brian Erard; and Jonathan Feinstein. "Tax Compliance," *J. Economic Literature* 36: pp 818–860, 1998.

Becker, Gary S. "Crime and Punishment: An Economics Approach," *J. Political Economy* 76: 169–217, 1968.

Boylan, Scott J. and Geoffrey B. Sprinkle, "Experimental Evidence on the Relation Between Tax Rates and Compliance: The Effect of Earned vs Endowed Income," *J American Taxation Association* 23:1 75–?? Spring 2001.

Chu, C, Y. C. (1990): "Plea Bargaining with the IRS." *Journal of Public Economics* 41: 319–333.

Cowell, F. A, (1990): "Tax Sheltering and the Cost of Evasion." *Oxford Economic Papers* 42: 231–243.

Falkinger, Josef, and Herbert Walther. "Separating Small and Big Fish: The Case of Income Tax Evasion," *Journal of Economics* 54:1 55–67, 1991.

Feld, Lars P., and Bruno Frey. "Trust Breeds Trust: How Taxpayers Are Treated Institute for Empirical Research in Economics," University of Zurich Working Paper No. 98, January 2002.

Lazear, Edward P.; Ulrike Malmendier; and Roberto A. Weber. "Sorting in Experiments with Application to Social Preferences," *American Economic Journal: Applied Economics* 4:1, 136–163, 2012.

Phillips, Mark D. "Reconsidering the Deterrence Paradigm of Tax Compliance." 2011 IRS-Tax Policy Center Research Conference: New Perspectives on Tax Administration, June 22, 2011. Accessed at http://www.irs.

gov/uac/SOI-Tax-Stats---2011-IRS-TPC-Research-Conference.

Posner, Eric A. "Law and Social Norms: The Case of Tax Compliance," *Virginia Law Review* 86: 1781–1819, 2000.

Raskolnikov, Alex. "Revealing Choices: Using Taxpayer Choice to Target Tax Enforcement," *Columbia Law Review* 109: pp 689–754, 2009.

Slemrod, Joel, Marsha Blumenthal, and Charles Christian, "Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota," *J. Pub. Econ.* 79: 455, 465 (2001).

World Values Survey. (2011) *WVS 2010–2012 Questionnaire, Revised Oct, 2011*. Available at http://www.worldvaluessurvey.org/wvs/articles/folder_published/article_base_136/files/WVS_2010-2012_REVISED_OCT_2011.pdf.

Yitzhaki, Shlomo. "A Note on Income Tax Evasion: A Theoretical Analysis." *J. Public Economics* 3: 201, 1974.

Zelenak, Lawrence. "Tax Enforcement for Gamers: High Penalties or Strict Disclosure Rules?" *Columbia Law Review,* Sidebar: Response to Alex Raskolnikov. 24 June 2009.

Endnotes

1 Examples of such programs include the Tip Rate Determination Agreement (TRDA), Tip Reporting Alternative Commitment (TRAC), and the recently discontinued Attributed Tip Income Program (ATIP).

2 In this analysis, we assume that taxpayers continue to file required returns.

3 Unless, as Raskolnikov suggests, the authority overrules their choice, assigning CR gamers identified by audit into DR and removing regime choice from a subgroup of taxpayers. As Zelenak (2009) points out, this is inconsistent with rationalizing a two-regime proposal on the basis of taxpayer choice.

4 A description of the Social and Behavioral Sciences Lab (SBSL) is available at http://sbsl.umn.edu/.

5 Boylan and Sprinkle (2001) demonstrate that experimental subjects are likely to respond differently to an increased tax rate when they are simply endowed with income, relative to when they earn it. In order to capture more closely their real-world situations, we elected to required subjects to earn income.

6 World Values Survey (2011), Question V201, p. 14.

7 U.S. Bureau of Labor Statistics, Labor Force Statistics from the Current Population Survey, Series id LNS12300000,(Seas) Employment-Population Ratio, available at http://data.bls.gov/timeseries/LNS12300000.

8 U.S. Bureau of Labor Statistics, Household Data Annual Averages, Table 10: Employed Persons by occupation, available at http://www.bls.gov/cps/cpsaat10.pdf.

9 The Likert scale for the World Values Survey question is the reverse of the scale for our question. Reversing the scale to compare responses, our subjects' mean response was 3.8, which is larger than the mean response for U.S. World Values Survey respondents (2.1). See http://www.wvsevsdb.com/wvs/WVSAnalizeQuestion.jsp.

10 We ran two-limit Tobit regressions, controlling for subjects who did not underreport any income as well as subjects who underreported all of their income.

11 Although we label these indices "propensity-to-cheat," some observed underreporting could be due to unintentional error.

12 Coefficients on all of the demographic control variables are available from the authors.

13 Coefficients on all of the demographic control variables are available from the authors.

14 These are two-limit Tobit regressions, controlling both for subjects who had zero underreported income and for subjects who underreported all of their income.