Behavioral effects of tax withholding on tax compliance: Implications for information initiatives

Christian A. Vossler[a,*], Michael McKee[b], and David M. Bruner[b]

[a] Department of Economics and Howard H. Baker Jr. Center for Public Policy, University of Tennessee, Knoxville, TN 37996
[b] Department of Economics, Appalachian State University, Boone, NC 28608

The published version of this study is in the *Journal of Economic Behavior and Organization*. The final publication is available at https://doi.org/10.1016/j.jebo.2020.12.030.

© 2020. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/.

* Corresponding author: Correspondence should be directed to Christian Vossler, Department of Economics, 527F Stokely Management Center, University of Tennessee, Knoxville, TN 37996. E-mail: cvossler@utk.edu. Telephone: 865-974-1699. Fax: 865-974-4601. This research has benefited from several discussions with Kim Bloomquist and Alan Plumley at the IRS. Jens Schubert, Nathan Murray, and Adrienne Sudbury provided exemplary research assistance. We thank Daniela Puzzello and two anonymous reviewers for their insightful comments and suggestions.

Behavioral effects of tax withholding on tax compliance: Implications for information initiatives

Abstract

Using a theory-driven experiment with working adult participants and deliberate tax framing, this paper reports on the effects of tax withholding choices on subsequent individual income tax reporting behavior. We find reporting increases with the level of taxes over-withheld, and decreases with taxes under-withheld, with the latter effect being much larger in magnitude. We explore two information programs designed to influence social norms: compliance information specific to one's income class, and information related to fairness of the tax system. Both increase tax reporting but only information on fairness affects the withholding choice. A third information program – a service that resolves uncertainty over tax liability – offsets the undesirable effects of tax under-withholding on reported liability when acquired. Using information from an extensive questionnaire, we find several interesting associations between decision-making in the experiment and the prior tax filing experiences and behaviors of participants.

JEL Classifications: H21, H26, C91, C92

Keywords: tax withholding; tax assistance services; social norms; fairness; tax reporting and enforcement; experimental methods; framed field experiment

1. Introduction

A central feature of many individual income tax systems is that people are required to withhold, or pay estimated, taxes during the tax year; for example, employers withhold taxes before wages or salaries are paid to employees.[1] Such withholding represents a substantial change to the tax reporting decision as it has been framed in the literature. The standard theoretical model (see Allingham and Sandmo, 1972; Yitzhaki, 1974) adapts the "economics of crime" approach pioneered by Becker (1968) and evaluates tax evasion as a "gamble" in which the default position is zero tax reported. Tax withholding changes this default and, in effect, makes the compliance decision a two-step process.

If final liabilities are known with certainty when the withholding decision is made, the individual can choose, via the withholding decision, the baseline for compliance. However, liabilities are often uncertain at the withholding stage, and withholding calculation methods are imprecise, giving rise to the possibility that an individual is in an unexpected tax owed or tax refund due situation at the end of the tax year. Compounding this is the fact that those in a tax owed situation may face additional costs due to a liquidity constraint and/or penalties for withholding less than the reported liability.[2] The result is an asymmetric decision setting when filing, akin to the zero owed/refund position being an induced reference point. Thus, withholding introduces scope for behavioral effects in the reporting decision.

We use theory and an incentivized policy experiment, with working adults as participants, to gain new insight on the role of withholding on subsequent tax reporting behavior.

[1] Withholding at the source has been in effect in the U.S. since World War II, and has generally increased revenue. From 2008 to 2010, the U.S. IRS (2016) estimated compliance is higher when income is subject to withholding.
[2] For example, the U.S. IRS imposes under-withholding penalties, and reports that 10 million taxpayers paid these in the 2015 tax year (U.S. IRS, n.d.). Avoiding such penalties is the focus of Gandhi and Kuehlwein (2014).

Within this framework we examine the role of tax agency information programs that can resolve tax liability uncertainty or provide information on peer behavior. Information programs have the potential to directly influence reporting decisions, and indirectly affect reporting by altering the withholding choice. We induce uncertainty over liability in both the withholding and reporting stages, which can reflect tax code complexity, and imperfect recording keeping on sources of income and deductions, such as tip income and charitable contributions.

Whereas there is now a sizable literature utilizing laboratory experiments to study tax compliance (Alm, 2019), most studies have abstracted away from withholding. As exceptions, some authors have used hypothetical choice scenarios, yielding a common but not universal finding that those in an over-withholding position are less likely to evade.[3] Intrinsic considerations can dominate in hypothetical choice settings, and may give rise to behaviors that differ from field settings with salient financial incentives.

Allen (2018) employs an incentivized setting that addresses the link between real effort and tax withholding. This is an important issue but our focus is on tax reporting *after* earnings are realized rather than the effect of taxation on labor effort.[4] Economists have also used field experiments to study tax compliance (e.g., Blumenthal, Christian, and Slemrod, 2001; Hallsworth *et al*., 2017; Kleven *et al*., 2011; Slemrod *et.al*., 2017), but this research has not focused on the role of withholding on compliance.

A handful of studies using observational data provide insight on the links between withholding and reporting. Using field data from the Taxpayer Compliance and Measurement

[3] See Copeland and Cuccia (2002), Kirchler and Maciejovsky (2001), Martinez-Vazquez, Harwood, and Larkins (1992), Robben *et al*. (1990a), Robben *et al*. (1990b), Schadewald (1989), and Schepanski and Shearer (1995).
[4] Similarly, Becker, Fooken, and Steinhoff (2019) use an incentivized lab experiment to study withholding, demonstrating that higher withholding rates reduce effort (i.e., labor supply), while refunds increase effort. However, noncompliance with tax payments (the focus of our research) is not possible in their experimental setting.

Program, Clotfelter (1983) and Chang and Schultz (1990) find a positive correlation between withholding and reporting, with those who under-withhold more likely to underreport. More recently, Engstrom *et al*. (2015) use a regression kink and discontinuity approach to identify the effects of claiming a specific deduction in the 2006 tax year. Rees-Jones (2018) applies a bunching-based identification strategy to data from the IRS Panel of Individual Tax Returns. Results from these two studies strongly suggest there is a causal link – those in an under-withholding position report less, which the authors attribute to loss aversion. Carillo and Emran (2018) exploit changes in the withholding rates affecting firms in Ecuador, and find that in some industries firms increased tax payments when facing a higher withholding rate.

Relative to observational studies, regardless of whether an audit occurs, in our experiment we are able to know precisely both liability and evasion and further to disentangle the effects of being in an over- or under-withholding position. Relative to past laboratory experiments examining withholding, we use financial incentives, allow participants to select a withholding level, and elicit a continuous measure of compliance.[5] Our experimental setting implements the annual cycle with a withholding stage at the beginning of the tax year and a filing stage at the end. In an induced-value setting subjects earn income, select a withholding level (multinomial choice), and file a tax return by reporting liability (a continuous choice). To enhance external validity, we use deliberate tax framing featuring institutional details, such as the use of a tax form, and a choice between a standard and itemized deduction. As our participants are working adults with diverse tax filing experiences, this study may be classified a "framed field experiment". Last, we go beyond prior empirical work by examining the effects

[5] In prior experiments the participant's withholding status is based on random treatment assignment. One exception is Copeland and Cucia (2002), who base withholding position on the participant's prior tax returns. Most studies related to withholding have examined settings where the compliance choice is binary rather than continuous.

that three information programs have on withholding and compliance behavior.

We incentivize experiment participants to carefully consider the withholding choice through a withholding cost that captures (opportunity) costs associated with pre-payment of taxes and a penalty for under-withholding.[6] The design introduces uncertainty over liability, and varies the information about liability across the withholding and reporting stages. This feature reflects the realities of many taxpayers as changes in the tax code, the coarseness of withholding methods, and changes in income streams can leave taxpayers in an unexpected refund or payment situation when filing. While important for parallelism, the exogenous liability changes implemented also help us causally identify the effects of withholding on subsequent reporting.

Tax agencies, such as the US IRS, have devoted increasing budget allocations to taxpayer services such as information resources (Vossler and McKee, 2017) on the grounds that this increases compliance. We examine the interaction between services and withholding as a component of compliance since a taxpayer's attitude toward compliance is predicted to be affected by the net tax position (additional taxes owed versus refund due) at the time of filing.[7] In particular, we study an assistance service that resolves liability uncertainty, information on the compliance of others, and a fiscal exchange.[8] Resolving uncertainty over tax liability has the potential to substantially alter reporting behavior. Prior research also suggests that taxpayers are expected to be influenced by the fiscal exchange (the perceived public good benefits arising from taxes paid) and the reporting behavior of others.

---

[6] At the time of the experiment, the US Individual Income Tax form 1040 included a line requiring interest (and penalties) for tax underpayment (including under withholding) be included during the current tax year.

[7] Alm, Jackson and McKee (2009) use a lab experiment to study the transmission of audit enforcement information on compliance while Pomeranz (2015) studies compliance with the European value added tax (VAT), with results that support those of Slemrod, Blumenthal, and Christian (2001). Studying a different tax base than ours, Brockmeyer and Hernandez (2016) examine information on business taxation in Costa Rica. Slemrod, *et al.* (2017) study a new IRS reporting rule as a means of detecting potential business taxpayers who are not in the system.

[8] Compliance can be associated with "obeying rules" and there is prior research on motives for obedience. This is beyond the scope of this paper since identification would require additional treatments beyond those conducted.

The theory offers new insights on how withholding, when modelled as a choice, affects reporting. First, whereas several studies suggest that reporting differences due to withholding position can only be explained by prospect theory, we show that a similar asymmetry arises under expected utility when there are (opportunity) costs to withholding taxes, and penalties or borrowing costs associated with paying additional taxes when filing. A second implication of our model is that, in contrast to a model that either ignores withholding or assumes it is costless, when withholding is costly it becomes optimal to report less in taxes.

In the experiment, reporting increases with the amount withheld, but increasing withholding beyond one's expected liability has a smaller effect than does increasing withholding for those in an under-withholding position. Although we confirm the basic result found elsewhere in the literature, we have thus found that the magnitude of the under-withholding level is of further importance to compliance.[9] Our more novel finding is that the undesirable effects of reporting for those in an under-withholding position are offset when the agency provides the taxpayer with more precise liability information. This has important resource allocation implications for the tax agency. Perceived fairness of the fiscal exchange influences withholding but compliance norms do not.

The theory and data further show that other results reported in the literature continue to hold when withholding is added to the reporting decision setting: evasion decreases with enforcement effort, social norms of compliance, and perceived fairness of the fiscal exchange.[10] We also find that a service that resolves liability uncertainty increases compliance.[11] Last, we

[9] Fochmann and Wolf (2019) report evidence of similar asymmetry and attribute this behavior to framing effects.
[10] See Alm, Jackson, and McKee (1993), Cummings *et al*. (2009) and Besley *et al.* (2019) for evidence on the link between tax compliance and the perceptions of norms and the fiscal exchange.
[11] This result is consistent with Alm *et al.* (2010), McKee, Siladke, and Vossler (2018), and Vossler and McKee (2017), all which involve random audits. Vossler and Gilpatric (2018) study endogenous audits, and find a liability information service combined with a guarantee limiting costs of detected underreporting reduces underreporting.

find suggestive evidence that taxpayer characteristics and experiences outside the lab are tied to behavior. Risk-averse individuals, women and older persons are less prone to underreporting. Participants with greater evasion opportunities in the field report less taxes in the experiment.

2. Theoretical Framework

We develop a theoretical framework to help motivate and inform the experimental design, as well as to support testable hypotheses. First, we build upon theories of individual tax reporting that consider tax liability to be uncertain (e.g., Alm, 1988; Vossler and McKee, 2017) by incorporating withholding, and an under-withholding penalty. This model is then extended to consider reference dependence and loss aversion, as well as the effects of information programs. Last, we model as a two-stage problem the optimal withholding and reporting choices. Most of the prior theory literature abstracts away from withholding. As exceptions, Yaniv (1999) considers the effects of withholding on compliance but does not model withholding as a choice. Feltham and Paquette (2002) include withholding as a decision variable, but assume that liability is certain when filing, and that the tax agency always detects and penalizes underreporting.

The theoretical framework applies to sources of income and deductions that are not subject to third-party reporting requirements, such as tip income and charitable contributions.[12] In most tax systems, the reporting "gamble" largely applies to these sources. Importantly, taxpayers naturally have uncertainty over the tax liability associated with these "unmatched" sources, for instance due to bookkeeping errors, measurement error associated with estimating values for donated goods, and uncertainty over the tax code. Tax liability uncertainty as well as the lack of third-party reporting requirements leads to discretion over tax withholding.

[12] In their field experiment, Kleven *et al.* (2011) find nearly zero evasion on income subject to third-party reporting, but substantial evidence of evasion for self-reported sources of income.

We proceed by characterizing optimal reporting for any possible withholding amount. Then, using backward induction, we model the optimal withholding choice. At the reporting stage a taxpayer chooses what liability to report, denoted by R. From the perspective of the taxpayer, liability is a random variable x with a distribution function $F(x)$, which is assumed to have positive density $f(x)$ on the interval $[a, b]$.[13] Let x^0 denote the true liability, which is assumed to lie within the interval. Institutional or other constraints restrict reports to lie in the interval $\left[\underline{a}, \overline{b}\right]$, with $\underline{a} \leq a \leq b \leq \overline{b}$.

Let W denote the withholding choice, which is determined prior to filing. In the theory, we will refer to under- and over-withholding as *optimally* paying additional taxes or claiming a refund at filing, respectively. Those in an under-withholding position pay a penalty equal to $\varphi(R - W)$ if $R > W$, where $\varphi > 0$ is the penalty rate. This may be imposed by the tax authority, and/or φ may reflect other financial considerations such as borrowing costs.

Upon filing, the authority audits a return with probability p, and audits are completely random and independent of whether others are audited or the reported liability. Upon audit, all unpaid taxes are required to be paid and there is marginal penalty $\beta > 0$ assessed on unpaid taxes. Adopting an expected utility framework, and assuming risk-neutrality, the optimal reporting problem is one of maximizing the expected payoffs associated with reporting taxes:

$$[1] \qquad \max_R \ \left[W - R - 1_{[R>W]}\varphi(R - W)\right] - \left[p(\beta + 1)\int_R^b (x - R)f(x)dx\right],$$

where $1_{[R>W]}$ is an indicator that equals 1 when the taxpayer reports taxes above withholding, and otherwise equals 0. The first bracketed term reflects the payoffs from reporting, unconditional on being audited. This amount is positive when a refund is being claimed, and is

[13] There are several "decisions" that jointly determine the amount of taxes reported. We abstract from this complication.

otherwise negative. The second bracketed term is the expected audit cost. When it is optimal to claim a refund, the under-withholding penalty is irrelevant, and the first-order condition is:

[2] $$1 = p(\beta + 1)\big(1 - F(R^*)\big).$$

An interior solution exists for R^* on the interval $[a, b]$ if $p(\beta + 1) > 1$. Otherwise, there is a corner solution $R^* = \underline{a}$, i.e., the taxpayer engages in maximum evasion. In general, it is possible for the optimal reported liability to be under, over or equal to the true liability. For instance, even if $E[x] = x^0$ (i.e. beliefs are unbiased) there is potential value in over-reporting in expectation as it decreases the probability (and expected cost) of underreporting taxes.

When it is optimal to pay additional taxes upon filing, reporting is characterized by:

[3] $$1 + \varphi = p(\beta + 1)\big(1 - F(R^{**})\big).$$

Although the withholding choice does not enter into the first-order condition, comparing [3] to [2] makes it clear that the withholding choice does matter. In particular, noting that $1 - F(R)$ is decreasing in R, under-withholding induces a lower reported liability; i.e., $R^{**} < R^*$.

The under-withholding penalty introduces a discontinuity in the maximand, and for withholding between R^{**} and R^*, reported liability equals the amount withheld, $R = W$. Following from [2] and [3], within this region the expected marginal compliance cost is strictly between 1 and $1 + \varphi$. Thus, reporting a dollar more than withholding increases cost by $1 + \varphi$, which exceeds the decrease in marginal compliance cost. Reporting a dollar less than withholding is not beneficial as the dollar saved is less than the increase in expected compliance costs. Hence, assuming an interior solution, we can characterize optimal reporting as follows:

[4] $$R = \begin{cases} R^{**} & \text{if} \quad W < R^{**} \\ W & \text{if} \quad R^{**} \le W \le R^* \\ R^* & \text{if} \quad W > R^* \end{cases}$$

There is overall a (weakly) increasing relationship between withholding and reporting.

Note that both R^{**} and R^{*} are increasing in the audit probability and the audit penalty. Therefore, as the strength of the enforcement regime increases, the range of withholding amounts for which it is optimal to pay additional taxes increases. Moreover, the range of withholding values for which the optimal report equals withholding decreases.

2.1 Reference Dependence

Prior work on the relationship between withholding and compliance has emphasized the potential role of prospect theory (Kahneman and Tversky, 1979) in characterizing behavior. Here, we incorporate reference-dependent preferences in a manner that closely follows Engström *et al.* (2015) and Rees-Jones (2018). In particular, we assume that reporting zero additional taxes due at filing serves as a natural reference point, and that claiming a refund results in an additional "gain" while paying additional taxes is seen as a "loss". Assume gain utility is equal to $\eta(W - R)$ for $W > R$ and loss utility is equal to $-\eta\mu(R - W)$ for $R > W$, where $\eta > 0$ captures the weight placed on gain/loss utility relative to monetary payoffs and $\mu > 1$ captures loss aversion (i.e., losses loom larger than gains). Optimal reporting maximizes the sum of expected payoffs and gain/loss utility:

$$[5] \qquad \max_{R} \ \left[W - R - 1_{[R>W]}\varphi(R - W)\right] - \left[p(\beta + 1)\int_{R}^{b}(x - R)f(x)dx\right] + \eta\left[(1 - 1_{[R>W]})(W - R) - 1_{[R>W]}\mu(R - W)\right].$$

When it is optimal to claim a refund, the first-order condition is

$$[6] \qquad 1 + \eta = p(\beta + 1)\big(1 - F(R^{*})\big),$$

and when it is instead optimal to report additional taxes owed the relevant condition is

$$[7] \qquad 1 + \varphi + \eta\mu = p(\beta + 1)\big(1 - F(R^{**})\big).$$

Comparing [6] with [2] (or [7] with [3]) implies that reference dependence decreases

reporting. Those in a "gain" situation will be driven to claim a larger refund, and those in a "loss" situation will report less to avoid the disutility of the loss. Similar to the effects of the under-withholding penalty, reference dependence drives a wedge between optimal reporting for those in an over- versus under-withholding position. This difference increases with the weight placed on gain/loss utility and value of the loss aversion coefficient. The range of values for which the optimal report equals withholding also increases with these parameters.

2.2 The Effect of Social Norms and Fairness of Fiscal Exchange (Public Goods)

There is an extensive literature investigating the role social norms play in the compliance decision (Alm, 2019; Alm, McClelland, and Schulze, 1999; Hashimzade, Myles, and Tran-Nam, 2013; Torgler 2002; Traxler, 2010). This work suggests there are also *implicit* costs for deviating from a compliance norm (Elster, 1989). Let α denote an exogenous reference compliance rate, with $0 \leq \alpha \leq 1$, which is multiplied by liability to determine the social norm, i.e. αx. Assume an implicit cost, $s(R)$, is incurred for deviating from the reporting norm,

[8] $$s(R) = \lambda(\alpha x - R)^2$$

where $\lambda > 0$.[14] This implicit cost can be the result of social sanctions, such as ostracism, resulting in a lost stream of future benefits from exchange with members of one's group, or emotional dismay, as a result of guilt or shame (Erard and Feinstein, 1994). The quadratic specification in [8] assumes a deviation above or below the norm is penalized symmetrically.[15]

Next, we account for the public goods and services paid from collected taxes (Alm, Jackson, and McKee, 1993; Alm, McClelland, and Schulze, 1999). Such benefits (costs) create

[14] Hence, the norm is the level of compliance not the level of reporting. The norm may thus differ by income class.
[15] Deviations above or below the norm could be penalized asymmetrically. For example, if only deviations below the norm are penalized (Alm, McClelland, and Schulze, 1999), higher compliance is induced.

an incentive for increased compliance (evasion). Likewise, considerations for fairness (Fehr and Schmidt, 1999) are important. Norms differ from fairness considerations in tax morale – the latter captures the perception of the tax burden, while the former addresses how others perceive the taxpayer's level of compliance. Hence, a taxpayer with high (low) morale employed in an industry with a norm of evasion (compliance) may be conflicted. Assume taxpayer j holds a subjective perception, $\pi(R_j)$, over the impact her report has on her own well-being, a tax morale. In general, these perceptions can either be a benefit, or a cost, and may either be increasing or decreasing in reported taxes. For tractability, let tax morale take the form of a public good,

$$[9] \qquad \pi(R_j) = \gamma \sum_{i=1}^{N} R_i - \delta\left(R_j - \bar{R}\right)^2$$

where $i = 1, \ldots, N$ is an index of taxpayers and $\bar{R}$ is the average contribution to the public good.[16,17] The parameter γ denotes the marginal per capita return from the provision of public goods and services, which may be perceived to be positive or negative (i.e. a public bad). Consistent with the experiment, we assume $1 > \gamma > 0$ such that paying taxes confers benefits, however, the marginal benefit is insufficient to induce voluntary provision of the public good (i.e. voluntary compliance) in the absence of other considerations. The quadratic term in [9] implies those who pay more (less) than average, and perceive that to be unfair, $\delta > 0$, incur a psychological cost and compensate by increasing evasion (compliance).

With the above behavioral considerations, the optimization problem expands to:

$$[10] \qquad \max_{R_j} \left[W - R_j - 1_{[R>W]}\varphi(R_j - W)\right] - \left[p(\beta+1)\int_{R_j}^{b}(x - R_j)f(x)dx\right] + \eta\left[(1 - 1_{[R>W]})(W - R_j) - 1_{[R>W]}\mu(R_j - W)\right] + \pi(R_j) - \int_a^b s(R_j)f(x)dx.$$

As before, interior solutions depend on whether it is optimal to claim a refund or pay additional

[16] In most settings, reporting will negligibly impact mean reporting of the group. Thus, we assume $\bar{R}$ is exogenous.
[17] Bordignon (1993) instead assumes fairness is a constraint, eliminating the possibility to tradeoff motives.

taxes at filing. In the first case, the optimal report is implicitly defined by:

$$[11] \qquad R^* = \frac{1}{2(\lambda+\delta)}\left[2\lambda\alpha E[x] + 2\delta\bar{R} + \gamma + p(1+\beta)\left(1 - F(R^*)\right) - 1 - \eta\right],$$

where $E[x]$ denotes the taxpayer's expected liability and for convenience we drop the subscript j. Hence optimal reporting becomes a weighted average of the social norm, $\alpha E[x]$, and the average contribution of all taxpayers, $\bar{R}$, according to their relative marginal costs of deviation, and is adjusted for the discounted effect of the expected marginal compliance cost and the marginal per capita return to the public good less the marginal cost of reporting.

If in an under-withholding position, $W < R^{**}$, optimal reporting is implicitly defined by:

$$[12] \qquad R^{**} = \frac{1}{2(\lambda+\delta)}\left[2\lambda\alpha E[x] + 2\delta\bar{R} + \gamma + p(1+\beta)\left(1 - F(R^{**})\right) - 1 - \varphi - \eta\mu\right].$$

The presence of the under-withholding penalty and, when applicable, loss aversion, continues to lower reporting when in an under-withholding position; i.e., $R^{**} < R^*$.

The experimental design varies the audit probability and introduces variation in the compliance norm and fairness measures. To derive how these factors alter optimal reporting, we can take derivatives with respect to [11] and [12]. For convenience, let $\hat{R}$ denote either R^* or R^{**}. We begin by examining the effect the audit probability has on reporting. Taking the derivative of [11] or [12] with respect to p yields

$$[13] \qquad \frac{\partial \hat{R}}{\partial p} = \frac{(1+\beta)\left(1-F(\hat{R})\right)}{[2(\lambda+\delta)+p(1+\beta)F'(\hat{R})]} > 0.$$

Reporting increases with the audit probability, as expected. Note that this (directional) result is robust to the presence or absence of behavioral motivations considered. Turning to the effect that the social norm, $\alpha E[x]$, and farness, $\bar{R}$, has on reporting:

$$[14] \qquad \frac{\partial \hat{R}}{\partial[\alpha E[x]]} = \frac{2\lambda}{[2(\lambda+\delta)+p(1+\beta)F'(\hat{R})]} > 0 \quad \text{and} \quad \frac{\partial \hat{R}}{\partial \bar{R}} = \frac{2\delta}{[2(\lambda+\delta)+p(1+\beta)F'(\hat{R})]} > 0.$$

Hence, reported liability is increasing in both the social norm and fairness measures.

2.3 The effect of a liability information service on tax reporting

Liability information services provided by the tax agency, which can take the form of walk-in sites, advice over the telephone, online calculators, and support documentation, have the potential to improve reporting accuracy. We examine a service that, if acquired by the taxpayer, eliminates uncertainty over liability. This service is analyzed theoretically by Vossler and McKee (2017), who show in their model the service induces reported liability that is closer to actual liability, on average. This remains true in our extended model.

As reducing uncertainty can either decrease or increase reported liability, here we establish some expectations based on the experimental design parameters. In the experiment, $F(x)$ is symmetric, which implies $F(E[x]) = 0.5$. $F(x)$ is also unbiased, as we present distributions used to determine actual liability. As a result, $E[x] = x^0$ on average. Last, we explore enforcement parameters such that $p(\beta + 1) \leq 2$.

We first establish that the service will increase reporting on average in the absence of social norm and fairness considerations. In the over-withholding setting characterized by [6] and imposing the symmetry assumption, as long as $p(\beta + 1) < 2(1 + \eta)$, it is optimal for someone in an over-withholding position to underreport liability on average; i.e., $R^* < E[x] = x^0$. With uncertainty eliminated, one weighs the cost of reporting another dollar, $1 + \eta$, with the benefits of reduced audit costs, $p(\beta + 1)$. These margins are independent of the reporting amount and thus, one either fully evades or fully complies. As long $p(\beta + 1) > 1 + \eta$ it is optimal to fully comply. Thus, when $1 + \eta < p(\beta + 1) < 2(1 + \eta)$, the service increases reporting. Otherwise, when $p(\beta + 1) < 1 + \eta$, it is optimal to engage in maximal evasion regardless of the service. Recalling that taxpayers in an exact or under-withholding position report less than R^*, it follows that the service (weakly) increases reported liability regardless of the withholding position.

To the extent social norms and fiscal fairness considerations matter, the service can either increase or decrease reporting on average. This is because the "cost" of deviating from social norms or the fairness criterion may be sufficient to induce over-reporting relative to expected liability. When it is optimal to claim a refund, with certain liability the solution to [10] is

[15] $R^* = \frac{1}{2(\lambda+\delta)}\left[2\lambda\alpha x^0 + 2\delta\bar{R} + \gamma + p(1+\beta)1_{[R^*<x^0]} - 1 - \eta\right]$.

Noting that $E[x] = x^0$, this solution differs from the uncertainty case, [11], only in terms of the differential weight placed on the expected marginal cost of an audit. When the solution under uncertainty is $R^* < E[x]$, then the liability service will increase reporting on average as $1 - F(R^*)$ will be less than 1. However, when $b > R^* > E[x]$, then obtaining the service will decrease average reporting. In this range audit penalties can still occur. With certainty, it is optimal to report at or above x^0, and the marginal audit cost is zero. When $\bar{b} > R^* > b$, audits result in no penalty, and resolving uncertainty will not alter average reporting.

2.4 Tax withholding

When choosing withholding, it is possible the choice will be based on planned reporting behavior (i.e., taxpayers will use backward induction). Withholding is potentially beneficial as it reduces costs associated with the under-withholding penalty. Given taxes withheld are fully credited when filing, if there is no cost of withholding one should voluntarily withhold as much as possible. At the other extreme, if withholding costs exceed the under-withholding penalty at the margin, then no taxes should be voluntarily withheld. We focus on the intermediate case, letting $c > 0$ denote the marginal (opportunity) cost of withholding, and assume $c < \varphi$.

First consider a situation where withholding is an unconstrained, continuous choice and the information set in the withholding stage is identical to the information set when filing. In this

case, the optimal withholding choice equals the optimal reporting choice. We established above optimal reporting equals withholding as long as $R^{**} \leq W \leq R^{*}$. Thus, to support the claim we need to show other withholding choices are suboptimal. For any $W < R^{**}$, it is optimal to report R^{**}. As taxes are under-withheld, there is an under-withholding penalty, and since this (marginal) penalty is greater than (marginal) withholding cost, money is saved by withholding at least $W = R^{**}$. When $W > R^{*}$, taxes are over-withheld, and a refund is claimed. Excess withholding costs are avoided by reducing withholding to at least $W = R^{*}$.

The optimal withholding choice minimizes the sum of the expected costs across both the withholding and reporting stages. For illustration, and using the fact that $W = R$, consider the optimal withholding choice based on the reporting model [1]:

[16] $\min_{W} \; [W(1+c)] + \left[p(\beta+1)\int_{W}^{b}(x-W)f(x)dx\right].$

The first-order condition is:

[17] $1 + c = p(\beta+1)\big(1 - F(W^{*})\big),$

which implicitly defines the optimal withholding (reporting) choice. With $c > 0$ arbitrarily small, this yields $W = R^{*}$ as the optimum (i.e., [17] is identical to [2]). With c sufficiently close to φ, one withholds $W = R^{**}$ ([17] is equivalent to [3]). For other values, optimal withholding (reporting) falls between the two amounts. One implication is, in contrast to a model that either ignores withholding or assumes it is costless, when withholding is costly this lowers optimal reporting.

Next, consider the case where liability information can differ between the withholding and reporting stages. For instance, in the withholding stage a taxpayer may envision two possible reporting scenarios, one corresponding with “high” earnings and another with “low” earnings. When filing, whether earnings were “high” or “low” is known. If the taxpayer is forward looking

and optimally selects withholding, she needs to first determine how, conditional on some withholding amount, the optimal report for each possible liability scenario in the reporting stage. Let $V(\hat{R}(W), W, F(x))$ denote the value function associated with the optimal reporting problem (e.g., [1] evaluated at $\hat{R}$, the solution to the problem), which defines the maximal reporting benefits conditional on the withholding choice, and the assumed liability distribution. Then, the optimal withholding problem is:

[18] $\min_{W} \; [W(1+c)] - \sum_{k=1}^{K} r_k V_k(\hat{R}_k(W), W, F_k(x))$,

where r_k denotes the probability a scenario (e.g., "high" earnings) arises. The first-order condition is:

[19] $1 + c = \sum_{k=1}^{K} r_k \frac{dV_k(\hat{R}(W), W, F(x))}{dW}$.

The taxpayer now balances the marginal withholding costs incurred in the withholding stage with marginal benefits of withholding to be realized, in expectation, in the reporting stage. In particular, as optimal reporting in each scenario is (weakly) increasing in withholding (by either zero or one dollar, as implied by [4]), withholding increases the refund payment, and decreases the under-withholding penalty and audit costs. As a single withholding amount is selected, it is not possible for this choice to be optimal for every possible liability scenario. This means optimal withholding can result in an over- or under-withholding position depending on what scenario is realized prior to filing.[18] Further, as in the case without liability changes, the same factors that increase reporting will continue to increase optimal withholding. For instance, an audit rate increase increases reporting for every liability scenario, which increases withholding.

[18] In many tax systems, institutional rules create a constrained, discrete choice problem for withholding. When one must select from a set of possible withholdings, regardless of whether liability information changes from the withholding and reporting stages, additional cases arise where optimal withholding deviates from optimal reporting.

3. Experimental Design

3.1 Experiment Setting Details

The decision setting parallels key components of the field setting.[19, 20] Experiment parameters are reported in Table 1, with amounts denominated in lab dollars.[21] Participants earn income by performing a task, choose tax withholding, self-report their liability, and then face a possible audit and penalties for underreporting. In the earnings task, participants are presented with a picture of either a jar of pennies, gumballs or jelly beans, and guess the number of elements in the jar. One-third of participants with the closest guess are placed in the highest income class, the second third in the middle income class, and the next third in the lowest income class. Income class assignments are in effect for a "series" of rounds. A new income task begins each series – thus participants may experience different income levels.

A decision round begins with participants making a withholding choice, through a form that mimics the W-4 form of the IRS.[22] There are five withholding levels to choose from (the number of "allowances" claimed). The withholding amounts span a wide range, allowing participants to unambiguously over-withhold (claim zero allowances) or under-withhold (claim four allowances) relative to liability. To inform this choice, information on liability is provided: a range of possible incomes; two possible standard deductions; and, a range of possible itemized deductions. The true income and itemized deductions lie within these ranges. This information allows one to determine expected liability, albeit with considerable uncertainty.[23] The cost of

[19] Sample subject computer screens and printed instructions are provided in Appendix A.
[20] As our objective is to inform policy, the experiment incorporates many features of the naturally occurring setting. This allows us to investigate many nuances of the interaction between withholding and reporting in a setting where final liability is often uncertain. Offsetting the experiment complexity is the fact that our subject pool consists entirely of working adults who have filed tax returns in the past and a relatively large sample size.
[21] Lab dollars are converted to US dollars at the end of the session at the rate of 300 lab dollars to one US dollar.
[22] A copy of the W4 form in effect at the time these sessions were conducted is included in Appendix A.
[23] We do not allow for revisions to withholding during the decision round. Importantly, Jones (2012) provides evidence that taxpayers are slow to adjust withholding in response to changes in their own status or tax policy.

withholding is equal to the amount withheld plus an additional cost of 10%, which reflects factors such as opportunity cost.[24]

After the withholding choice, participants are provided, after a software implemented time delay, with a tax form. Participants report income and claim either a standard deduction (there are two possible amounts designated "high" and "low") or an itemized deduction (an amount of their choosing). Final liability is the difference between income and deductions claimed (i.e., taxable income) multiplied by a tax rate of 50%. Recall that ambiguity in the tax code as well as the potential for unanticipated income changes during the year contributes to liability uncertainty when withholding. Some, but not all, of this uncertainty is likely to be resolved over time. Relative to the withholding stage, the actual income and itemized deduction ranges are randomly and independently shifted up, down, or centered on the initial distributions. Further, these ranges are reduced by 25%.[25] Also, the true standard deduction amount is revealed. These processes serve to hold fixed, increase or decrease expected liability across the withholding and reporting stages. This exogenous variation helps to causally identify the effects of withholding on reporting.

Participants can alter their tax form entries until they file or until the form times out.[26] At any point they can update the form by clicking a "Do the Math" button whereby the amount of reported taxes is calculated along with the corresponding payment or refund. The amount of withholding is automatically credited on the form. Thus, a payment (refund) is revealed if reported liability is greater (less) than withholding. In the event that a payment is due, an under-

[24] The form is time limited, and failure to submit on time is equivalent to reporting no exemptions and results in the maximum level of withholding. This is the default value for anyone not submitting a W-4 to their employer.
[25] The 25% reduction in uncertainty was chosen to reflect a salient partial resolution while maintaining substantial uncertainty at filing, which preserves the value of the information service.
[26] Timing out results in an audit and, since entries are imputed to be zero, no deduction is claimed. Instructions informed participants it is not in their interest to time out. The form timed out in 0.3% of observations.

withholding penalty is also reported, equal to 20% of the reported payment.

Following the reporting stage, there is a randomly determined audit process.[27] The audit probability is known and audits are perfectly revealing. If audited, unpaid taxes (based on the actual income and allowable deduction amounts) are discovered and collected along with the penalty, equal to 300% of unpaid taxes. Following the audit process, a summary screen reveals the actual income and deduction amounts (i.e., random draws from the ranges provided in the reporting stage), what was reported, and details relevant for earnings calculations. Earnings are determined as the difference between actual income and taxes paid with withholding cost, under-withholding penalty, and audit costs subtracted from earnings, as applicable.

3.2 Experiment Treatments and Sessions

There are four between-subject treatments (Table 2). Treatment variables are the presence/absence of a liability information service and the presence/absence of information on compliance and a partial redistribution of taxes (i.e., a public good).[28] In T3 and T4, liability information assistance is offered in the reporting stage. The service resolves all liability uncertainty – the ranges of possible income and itemized deduction amounts are replaced by actual amounts.[29] To reflect transaction costs associated with obtaining information there is a monetized cost for the service, equal to 50 lab dollars.

When information on compliance and a public good is provided (T2 and T4), only the taxes *voluntarily* reported are used to finance the good. Implicitly this treats the penalties and

[27] Currently the IRS audit selection is largely based on endogenous rules, but the use of random audits allows us to focus on the research objectives. Although the evidence is dated, US taxpayer surveys suggest that many perceive audits to be random, and that audit rates are very high (Louis Harris and Associates, Inc., 1988).
[28] In previous research we systematically turned on/off the tax compliance and public good elements in the design (see Vossler *et al*., 2012). As we found no interaction effects, here we turn on/off both features simultaneously.
[29] Incomplete and/or incorrect services have been studied (see Alm *et al.,* 2010, and Vossler and McKee, 2017).

unpaid taxes collected as the cost of the audit process. The public good multiplier is set low since this is a tax reporting exercise not a public good provision game. Specifically, 50% of taxes voluntarily paid are equally allocated to all group members. We provide end-of-round information on the taxes reported relative to taxes owed, broken down by income class. Reporting can be compared to the income class average, and that of other income classes.

An experimental session consists of 18 paid rounds arranged into three series of six rounds each. Income and the audit rate vary across, but not within, these series. Three audit rates are used: 10%, 30% and 50%, and a participant faces each rate during the experiment. Within a session there are two distinct groups each consisting of three income classes. To control for order effects, there are six unique audit rate sequences, and the sequence is randomly varied across sessions, and across groups within a session. In the social norm/public good treatments, compliance information and transfer payments are group-specific. Group member are anonymous and only aggregate information regarding group behavior as it applies to the size of the public good and/or the overall compliance rate within the group is provided.[30]

3.3 Testable hypotheses

The experiment data allow us to test several hypotheses related to tax withholding and reporting, which we formally state below. All hypotheses follow directly from the theory.

Hypothesis 1. Reported liability increases with withholding (see equation [4]).

Hypothesis 2. Uptake of the liability information service increases reported liability ([15]).

[30] There are 18 unique sessions in the experimental design, the distinguishing features of which are presented in Table A.1. Given the interdependencies created by implementing the compliance/public good features, there are six sessions for each of the two treatments with these features. For the remaining two treatments there are three sessions each. In refining instructions, procedures, and design, we conducted three pilot sessions with student subjects.

Hypothesis 3. Withholding and reported liability increase with the compliance norm ([14]).

Hypothesis 4. Withholding and reported liability increase with average contributions to the public good ([11]).

Hypothesis 5. Withholding and reported liability increase with the audit rate ([13]).

Table 3 presents point predictions of tax reporting, conditional on withholding and expected liability in the reporting stage. Details on how the theoretical framework can be applied to generate these and other point predictions are provided in Appendix B. The withholding amounts reflect those available to a particular income group. Consistent with the above hypotheses, reported liability increases with withholding and the audit rate. At the 10% audit rate, theory predicts maximal evasion. The 30% audit rate motivates a modest amount of evasion. With a 50% audit rate and sufficient withholding, it is optimal to report one's expected liability. With lower amounts withheld, however, evasion is predicted.

Forward-looking taxpayers may optimally withhold based on the possible liability scenarios that can arise in the reporting stage. For illustrative purpose, as indicated by shaded table entries, we generated predictions for optimal withholding and reporting in cases where expected liability does not differ between the two decision stages. For the 10% audit rate, it will be optimal to select the lowest withholding amount, and those in the middle- and high-income classes should claim a refund when filing. At the 30% and 50% audit rates, optimal withholding is below and above, respectively, expected liability. The experiment is designed so that in most cases expected liability will differ between the withholding and reporting stages. In turn, even perfect optimizers will either be in an optimal refund due or taxes owed scenario when filing.

As it is optimal to fully evade with the 10% audit rate, regardless of withholding position, a liability information service that fully resolves uncertainty has no impact on optimal

reporting.[31] On the other hand, with a 30% or 50% audit rate, theory predicts truthful reporting when liability is certain. The liability service, if acquired, should therefore increase reported taxes for the 30% audit case regardless of withholding position, and increase reporting for the 50% audit rate for those who optimally report additional taxes owed. For those who optimally claim a refund, the 50% audit rate is predicted to induce people to report truthfully, but only in expectation. However, knowing the true liability means that the taxpayer can avoid paying excessive taxes (when the truth is below expected value) and avoid penalties (when the truth is higher than expected value). Therefore, the information service still has value. To provide insight on whether the taxpayer should acquire the service, we can use the theory model to calculate the difference in expected payoffs with and without information.

Based on our parameters, when it is optimal to claim a refund, the cost difference ranges from 31.25 to 46.89 for the 30% audit rate, and from 93.75 to 140.63 for the 50% audit rate. Theory predicts that people will pay 50 lab dollars for the service only at the 50% audit rate; although, risk averse people may buy at the 30% audit rate. For those who optimally report additional taxes owed, the value of the service is diminished considerably – while it remains optimal to increase reporting, this comes at an additional cost in the form of the under-withholding penalty. The value of the service is thus decreasing in the withholding amount.

3.4 Participant Pools and Procedures

Participants were employed (full and part-time) adults from the Knoxville, TN and Boone, NC areas. The labs are located at the University of Tennessee and Appalachian State University. Both include two-dozen networked computers, a server, and software designed for

[31] This result will not necessarily hold, however, when social norms and fairness considerations are sufficient to induce additional tax reporting.

this series of experiments. Recruiting was via the Online Recruiting System for Experimental Economics (ORSEE) developed by Greiner (2015). Participant pools were built using posters and email announcements to various community groups in each location. Registered persons were invited to a session via email, and permitted to participate in only one session. No participant has prior experience in this specific experimental setting. There are 359 participants. The number of participants by session and lab location are presented in Table A.1 (Appendix).

The experiment was conducted using z-Tree (Fischbacher, 2007), and identical procedures were used in both study sites. In a session each participant sits at a randomly assigned computer located in a cubicle, and is not allowed to communicate with other participants. Decisions are private and anonymous. Prior to the tax experiment, on-screen instructions guide participants through a risk multiple price list modeled after Holt and Laury (2002), as amended by Bruner (2012). Printed instructions for the tax experiment are provided to participants and read aloud by a moderator to ensure both common knowledge and that participants at each site received identical instructions. One practice round is conducted with the timers turned off and with the experimenter directing the participants on the use of the interface. A second practice round is then conducted with the timers running, with clarification questions addressed at its conclusion. Neither practice round affected earnings in the experiment.

The experiment proceeds for 18 paid decision rounds (neither the actual number of rounds or length of a series is pre-announced). After the final decision round, participants learn of their cumulative earnings from both the risk elicitation exercise and the tax experiment and are then directed to complete both a demographic and taxpayer attitude debriefing questionnaire. The demographic questionnaire elicits information on personal characteristics as well as tax filing experiences. The attitude questionnaire is adapted from Kirchler and Wahl (2010).

Participants are then called to the front of the room individually and paid their earnings in cash. Average earnings were approximately $80, and sessions averaged two hours.

4. Results

Table 4 provides information on our sample[32] and data. On average participants report 566 lab dollars. We can compare this with a measure of actual liability, in particular the taxes that would be revealed by an audit. This measure averages 689, suggesting 82% of taxes are reported on average. Throughout the analysis we characterize under-withholding (over-withholding), at the time of filing, based on whether withholding is lower (higher) than expected liability. Important for identification, there is substantial variation in withholding positions. In 34% of cases people under-withhold, and do so by 271 lab dollars on average. Reporting when in this position is substantially lower, at 466 lab dollars, which is 66% of actual liability. 65% of cases are in an over-withholding position when filing, over-withholding by 354 lab dollars. Among those in this position, reported taxes average 617 or 91% of actual liability. Scatterplots (Figure 1) show basic correlations between reporting and withholding, using participant-level data averaged across decision rounds. The top panel plots reporting and under-withholding, suggesting a negative relationship (Pearson correlation coefficient is -0.4131, with $p<0.01$). The bottom panel suggests that reporting increases with over-withholding, with the magnitude of this relationship being relatively smaller (Pearson correlation coefficient is -0.2310, with $p<0.01$).

[32] This is a diverse group. The average age is 38, and ranges from 18 to 68. 72% classify themselves as employed full-time and 23% as part-time employed. Types of employment cover a wide breadth, with 40% being in the education area, 10% in the food services sector, and 7% in retail trade. Average annual (individual) income is about $31,000, with considerable variation across the participants. The majority (89%) filed taxes for the prior tax year. There is also a fair amount of variation in underreporting opportunities with 50% self-reporting having non-wage income and 24% using itemized deductions. Of those identifying which form they used to file their prior return, approximately half report using the standard 1040 form. Consistent with the broader population a majority of our participants claimed a tax refund on their last return (75%). About one third used a professional tax preparer, and a very small fraction utilized IRS taxpayer services within the prior year.

About 40% of the time the liability information service is purchased when offered. Given the fee for this service, this reveals willingness to pay for the information. Those purchasing the service report 595 lab dollars on average, whereas those who do not purchase report 537 lab dollars. Reported taxes appear to be increasing with respect to the measures of compliance norm and fairness as defined in Section 2. When the compliance norm measure is above (below) average, 702 lab dollars (439 lab dollars) are reported on average. The same figure is 609 lab dollars (504 lab dollars), when the fairness measure is above (below) average.

As our policy-motivated experiment has several moving parts, the statistics presented above are just suggestive of possible relationships. To draw causal inferences, we now turn to an econometric analysis using the panel data. We employ linear regressions throughout, and include participant and round fixed effects to control for possible omitted factors.[33] Cluster-robust standard errors are computed in all regressions, allowing for valid inferences in the presence of heteroskedasticity and within-person serial correlation.

Inclusion of participant fixed effects means any covariate that does not vary across rounds of the experiment drops out of the regression. However, as we are interested in the effects of taxpayer characteristics (outside the lab) on reporting behavior (inside the lab), we estimate second-stage, cross-section regressions using the estimated fixed effects as the dependent variable and the decision round-invariant factors as explanatory variables. The OLS estimator for this second-stage model is consistent under the assumption that there is no linear correlation between the time-invariant factors and the fixed effects in the outcome model (see Wooldridge, 2010, pp. 358-359). Using estimated coefficients as the dependent variable of a regression

[33] As there are potential endogeneity concerns over including explanatory variables based on tax withholding or liability service purchase choices, we alternately estimated Models 1 – 5 using a System 2SLS FE estimator (see Wooldridge, 2005). For all models we fail to reject the hypothesis that these variables are exogenous. Estimation details and results are provided in Appendix C.

typically introduces heteroskedasticity (Saxonhouse, 1976), and we address this by calculating heteroskedasticity-robust standard errors for the second-stage regressions.

4.1 Tax reporting

Table 5 presents regression results for our main outcome variable, *Tax Reported*. Model 1 is the most parsimonious specification, and includes several variables that allow us to test the main hypotheses presented in Section 3.3. *Tax Withheld* is the withholding choice. The variables *Compliance Norm* and *Fairness* allow for differences in outcomes tied to the social norm and fiscal exchange treatments. Consonant with the theory, *Compliance Norm* equals the compliance rate in the prior round for those in the same income class (i.e., reporting divided by taxes owed), multiplied by expected liability. *Fairness* is calculated as average reporting in the prior round for the group. *Earned income*, a dummy for whether the high standard deduction is allowed (*High Standard*), and the expected itemized deduction (*Itemized*) allow for differences in reporting based on expected liability. *Audit Rate*, the audit probability, controls for enforcement effort.

The corresponding second-stage regression estimates, presented in the same column, include the participant characteristics defined in Table 5. In addition, we include the indicator *Liability Service Not Purchased* to capture any difference between reporting for those in the information treatment that did not purchase the information, and those in treatments where no liability information service was available. The indicator *Social Interactions* allows for the mean outcome to shift in treatments that include the social norm information and public good.[34]

Estimation results for Model 1 (Table 5) provide support for the main hypotheses. For

[34] These two indicators do not vary over decision rounds and thus fall out of the fixed effects regression. Nevertheless, they are potentially important in the second stage regression. *Compliance Norm* and *Fairness* are continuous variables and, when either are equal to zero (i.e., when taxes paid equal zero), we expect the mean outcome to differ considerably relative to treatments where these features are turned off.

each dollar withheld, participants report 18 cents more in taxes. Purchasing liability information services increases reporting by 101 lab dollars – a considerable effect. If unobserved tastes for evasion are driving the service purchase decision, we expect reporting to systematically vary between those who do not purchase the service when available and those for whom the service is not available, which does not appear to be the case as the coefficient on *Liability Service Not Purchased* is statistically insignificant. Reporting increases with the compliance norm and fairness measures. Reported liability increases with actual liability: more taxes are reported by those who have higher incomes, and less are reported by those with higher allowable deductions. Nevertheless, the estimated coefficients on the income and deduction variables are indicative of underreporting as, on average, participants report far less than 50 cents in additional taxes for each additional dollar in income. Consistent with theory, reporting increases with the audit rate.

For additional insight, Model 2 allows for the effects of withholding on reporting to differ depending on the withholding position. The coefficient on *Tax Under-withheld* suggests that participants report 37 cents less for every additional dollar under-withheld, a large effect. Those in an over-withholding position report 6 cents more for each dollar over-withheld. Table 3 predicts a strong relationship between withholding and reporting for those who withhold amounts less than their expected liability at filing. The predictions, however, suggest that reporting is invariant to withholding for those who over-withhold. Risk aversion provides one possible explanation for the observed relationship between over-withholding and reporting.[35]

Model 3 extends the specification by allowing the effects of the information initiatives to

[35] Allowing for risk aversion, the analog to [2] is $p + (1-p)\frac{u_0{}'}{u_1{}'} = p(\beta+1)\big(1 - F(R^*)\big)$, where $u_0{}'$ and $u_1{}'$ denote the marginal utilities associated with the payoffs in the no audit and audit states. Assuming decreasing marginal utility, $\frac{u_0{}'}{u_1{}'} < 1$, reporting increases with risk aversion. Moreover, the ratio $\frac{u_0{}'}{u_1{}'}$ decreases as withholding increases; therefore, reporting increases with withholding. A similar result arises if we assume reference-dependent preferences along with a "gain" utility function that is concave, rather than linear as assumed in the theory.

vary for those in an under-withholding position, and further for the effects of providing the liability service to vary depending on whether the social norm and fairness-related information initiatives are in effect. Being in an under-withholding position increases the impact of the liability service on reporting by almost 50%. This result can be reconciled with theory. As withholding decreases, theory predicts reporting decreases. As such, participants in an under-withholding position are more likely to report less than their expected liability with uncertainty which in turn increases the effect that the service has on reporting. On the other hand, the effects of the compliance norm and the fairness criterion do not depend on withholding position empirically. The relevant comparative statics ([14]) suggest any differences due to withholding depend on the marginal effect reporting has on the probability an audit reveals no additional taxes due, i.e., $F'(R)$. Since uniform distributions are used to induce liability uncertainty, $F'(R)$ is constant and the withholding position does not alter the effects of social norms and fairness.

The positive effects the liability service has on reporting decreases in the presence of social norms and fairness considerations. The service does, however, continue to increase reporting (e.g., by 37.18 dollars ($p < 0.05$) for those not in an under-withholding position). The direction of this interaction effect is predicted by theory, in part due to reporting scenarios arising where the service is predicted to reduce rather than increase reported taxes. Moreover, the social norm and fairness considerations diminish the reporting incentives generated by the audit process which, in turn, decreases the importance of resolving uncertainty on optimal reporting.

Across the three specifications we find significant correlations between tax reporting and participant characteristics. The coefficient on the variable *Risk Averse*, which is an indicator for risk-averse individuals based on data from the risk elicitation task, suggests risk aversion increases tax reporting – an anticipated effect. Females report more, a stylized fact from the

literature (Bruner, D'Attoma, and Steinmo, 2017), and the level of reporting increases with age. Joint-filers report more, consistent with the fact both parties are liable if evasion is detected. Those who report having non-matched income, and those who itemized deductions on their prior tax return, report significantly less. Those who had to pay additional taxes on their last return also report less. As expected, those who use professional tax preparation services report a higher liability in the experiment. A related result is those indicating they used tax advice from non-professionals report a higher liability as well. These results strongly suggest participants bring some "homegrown" tax filing experience to the lab and buttresses our claim that we have a framed field experiment. We summarize the main findings on tax reporting below.

Result 1. The main predictions of the theoretical framework are supported: reported taxes increase with withholding (Hypothesis 1), uptake of the liability information service (Hypothesis 2), the compliance norm (Hypothesis 3), the fairness of the fiscal system (Hypothesis 4), and the audit rate (Hypothesis 5).

Result 2. The effects of tax withholding on reporting are asymmetric, with stronger effects of under-withholding relative to over-withholding.

Result 3. The increase in tax reporting as a result of resolving uncertainty over liability is more pronounced for those in an under-withholding position, and lower when information related to social norms and fairness considerations is provided.

Result 4. Tax reporting is correlated with both participant demographics and tax reporting experiences. Those with greater evasion opportunities in the field underreport more in the experiment. Those who seek assistance from tax professionals or non-professionals in the field underreport less in the experiment.

4.2 Demand for liability information services

Table 6 presents linear probability models of liability information service purchases, using specifications that parallel Model 1 and 2 for reporting. Variables measured in lab dollars are scaled by 100. The propensity to purchase the service increases with withholding. However, when allowing for differential effects based on the extent of over- or under-withholding, this

effect is largely driven by a decreased propensity to purchase the service as under-withholding increases. This effect is modest, with the probability of uptake declining by about two percentage points for every 100 additional lab dollars under-withheld. In most cases, the liability service induces higher reporting, which involves paying additional under-withholding penalties (or incurring additional "loss" utility), decreasing the value of the service. Meanwhile, the effect of over-withholding is statistically zero, relaying that those in an exact withholding position are no more likely than those in an over-withholding position to obtain the service. This is intuitive as knowing true liability is less valuable if the prior is that no additional taxes are owed.

A strong driver of information demand is the audit rate. Increasing the audit rate by 10 percentage points increases service purchases by nearly 3 percentage points. As discussed in Section 3.3, the difference in reporting cost with and without information, given our experiment parameters, increases with the audit rate. This result is also intuitive, as the value of information should increase with the marginal cost of evasion.

Somewhat surprising is that expected liability has at best a marginal statistical effect. However, this may be explained by the fact that there is no relationship between the degree of uncertainty and taxable income in the experiment. The presence of $F'(R)$ in the comparative statistics ([14]) suggests that a marginal increase in either the fairness or compliance norms should increase reporting as uncertainty decreases. However, these effects are presumably small, which may explain why fairness does not have a discernable effect. The compliance norm has a marginally significant effect, but it is small: evaluated at mean expected liability, increasing the compliance rate by 50% has just a 3 percentage-point increase in the propensity to buy the service. Turning to effects of participant characteristics, on average, females are much less likely to purchase information (a 23 percentage-point difference), and those with a college degree are

more likely to acquire information (a 14 percentage-point difference).

Result 5. People in an under-withholding position are less likely to purchase the liability information service. Acquisition of the service increases with the audit rate and the compliance norm.

4.3 Tax withholding

As described above, there is considerable variation in withholding choices. The results from a regression of withholding is presented as Model 6 in Table 7. Theory predicts that the same factors expected to increase reporting should also increase withholding, and so the specification largely parallels that of the tax reporting regressions. Covariates related to expected liability are defined according to the information set available in the withholding stage. We also include an indicator for whether the participant was audited in the prior period.[36] We find that withholding is increasing in earned income in a manner that reflects that the financial incentives were strong: withholding increases by nearly 50 cents for every one-dollar increase in income, consonant with our 50% tax rate. Withholding is increasing with the audit rate, and there is a marginally significant and positive effect from being audited in the prior round. Both of these results suggest the audit regime is important in motivating withholding. Fairness matters but the compliance norm does not, which follows since participants were more likely to over-withhold relative to expected liability. Theoretically the compliance norm can induce taxpayers to report 50% of liability at most (i.e., full compliance), whereas the fairness norm has the potential to induce reporting more than 50% of liability. Hence, over-withholding is only consistent with fairness. The availability of a liability information service also has no impact on withholding. In the withholding stage the information service has yet to reveal its usefulness, especially to the

[36] Including this variable in the tax reporting or liability service acquisition regressions has no effect.

extent that uptake is related to being in an unexpected tax owed situation. Turning to the effects of participant characteristics, risk aversion, being female, and being older increases withholding while having non-matched income decreases withholding.

Result 6. Tax withholding increases with perceived fairness of the tax system, expected liability, and the strength of the audit regime.

5. Discussion

Although tax withholding by employers is a central component of the US individual income tax system, and the systems of most developed countries, there has been surprisingly little research conducted on its effects on reporting behavior, in particular in an experimental setting where actual liability is known. This study addresses the effect of the withholding choice on subsequent reporting and examines the interplay between taxpayer information initiatives, withholding, and reporting. Our results support the conjecture that the position at the time of filing – whether in a refund situation versus owing additional taxes – has a significant effect on reporting. Reporting increases with the level of over-withholding, and decreases with under-withholding, with the latter effect being much larger in magnitude.

Much of the prior laboratory experiment research on tax compliance has been silent on the withholding component. Nevertheless, through this study we demonstrate that several findings from the literature remain robust in our expanded setting: reporting increases as the audit rate, perceived fairness, and social compliance norms increase; females and risk averse persons report more. We further replicate, in our expanded setting, the result that an information service that resolves uncertainty over liability can increase reporting.

The new insight gleaned from this study suggests ways in which information-based programs can be targeted to increase compliance, whether provided by the tax agency or

otherwise. The liability information service we study resolves uncertainty over liability at the time of filing, and is shown to reduce evasion in the experiment. This effect is demonstrated to be stronger for those who withhold less than they actually owe. Thus, while the availability of this service may not influence withholding (as suggested by the data analysis) the service does appear to offset the effects of being in an under-withholding position. This is an important and policy relevant finding. The service has a smaller effect for those who have over-withheld, suggesting efficiency gains by targeting those that reveal or are otherwise suspected to have under-withheld (e.g., those receiving "1099 income" without taxes withheld).

Beyond this finding, our evidence suggests that reporting increases with the reported compliance of a reference group and with higher "fairness" or returns from the fiscal exchange. With respect to the former, one strategic approach would be to highlight a certain taxpayer group with a high compliance rate, such as an occupation where most workers only have matched income. Such targeted information would be more effective than highlighting overall compliance rates. With respect to the latter, our evidence suggests expanding signage emphasizing "tax dollars at work" is likely to promote tax morale, which can have both direct and indirect effects on evasion, since we find that increased perceptions over fairness also increase withholding.

Our results also suggest that information programs or other inducements that motivate taxpayers to make midyear withholding changes in response to unanticipated increases in income may be fruitful. Interestingly, a few years ago the state of North Carolina introduced a modified withholding structure (a new "NC4" form) that emphasizes penalties for under-withholding and accentuates uncertainty at the time of the withholding decision. Presumably this will motivate some to move to an over-withholding position. Of course, although our experiment does not speak to this directly, reminding taxpayers of the angst regarding making a tax payment or the

"insurance" afforded by planned savings may be effective in generally promoting withholding.

Our participant pool has considerable diversity in socio-demographic characteristics and tax filing experiences (outside of the lab). The extensive debriefing questionnaire provides several potential explanatory variables related to reporting. Results suggest information initiatives such as those described above may be more productive for males, younger persons, and those with significant opportunities to underreport (e.g., those with unmatched income, those who itemize deductions), as these factors are associated with lower reporting in the experiment.

An experimental study seeking to inform the policy debate must meet criteria for parallelism and external validity. Our design mimics the essential features of the individual income tax system. Evidence of external validity of this setting is provided by Alm, Blomquist and McKee (2015) who compare field data and laboratory data obtained via a similar design to our present setting. As with all empirical work, however, the results reported here are potentially limited to the present context. Still, our results comport with prior findings, where available (e.g., the effect of enforcement effort), suggesting our newer findings can inform the debate.

Meeting our objective of parallelism comes at the cost of complexity. While the majority of participants indicated good comprehension of the experiment instructions and had prior experience filing taxes, it remains an open empirical question as to whether similar results would have arisen from a less sophisticated design. For instance, similar to prior research, withholding could have been exogenously determined, simplifying decisions and enhancing identification (at the expense of parallelism). Requiring participants to choose withholding presents identification challenges. We address this by randomly varying expected liability between the withholding and reporting stages, and including participant and round fixed effects in the regression analysis.

Prior theoretical work that considers withholding on tax reporting has taken the

withholding choice as given. While we allow withholding to be a choice, the tensions we introduce through withholding cost and under-withholding penalty parameters only capture some of the incentives for withholding. Of course, withholding decisions in the field are driven by factors not considered here. Fruitful future work could examine the role of foregone consumption or investment opportunities during the tax year, liquidity constraints, and opportunities for revising withholding. These and other simplifications in our theory and experimental design could be relaxed to gain additional insight on this challenging but important decision setting.

6. References

Allen, Stuart. 2018. "Income Tax Withholding and Reference-Dependence: Experimental Evidence on the Role of Salience." SSRN working paper 316100.

Allingham, Michael G. and Agnar Sandmo. 1972. "Income tax evasion: A theoretical analysis." *Journal of Public Economics,* 1(4): 323-338.

Alm, James. 1988. "Uncertain Tax Policies, Individual Behavior, and Welfare." *American Economic Review*, 78(1): 237-45.

Alm, James. 2019. "What Motivates Tax Compliance?" *Journal of Economics Surveys*, 33(2): 353-388.

Alm, James, Kim Bloomquist, and Michael McKee. 2015. "On the external validity of laboratory tax compliance experiments." *Economic Inquiry*, 53: 1170-1186.

Alm, James, Todd L. Cherry, Michael Jones, and Michael McKee. 2010. "Taxpayer information assistance services and tax compliance behavior." *Journal of Economic Psychology* 31: 577-586.

Alm, James, Betty R. Jackson, and Michael McKee. 1993. "Fiscal exchange, collective decision institutions, and tax compliance." *Journal of Economic Behavior and Organization*, 22(3): 285-303.

Alm, James, Betty R. Jackson, and Michael McKee. 2009. "Getting the Word Out Enforcement Information Dissemination and Tax Compliance Behavior," *Journal of Public Economics*, 93(3-4): 392-402

Alm, James, Gary H. McClelland and William D. Schulze. 1999. "Changing the social norm of tax compliance." *Kyklos*, 52(2): 141-171.

Becker, Gary S. 1968. “Crime and punishment: An economic approach.” *Journal of Political Economy*, 76(2): 169-217.

Becker, Johannes, Jonas Fooken, and Melanie Steinhoff. 2019. Behavioral Effects of Withholding Taxes on Labor Supply. *Scandinavian Journal of Economics* 121(4): 1417-1440.

Besley, Timothy, Anders Jensen, and Thorsten Persson. 2019. “Norms, Enforcement, and Tax Evasion,” NBER Working Paper 25575 (February, 2019)

Blumenthal, Marsha, Charles Christian, and Joel Slemrod. 2001. “Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota.” *National Tax Journal*, 54(1): 125-138.

Bordignon, Massimo. 1993. “A fairness approach to income tax evasion.” *Journal of Public Economics*, 52 (3): 345-362.

Bruner, David M. 2012. “Multiple switching behaviour in multiple price lists.” *Applied Economics Letters*, 18(5): 417-420.

Bruner, David M., D'Attoma, John, and Sven Steinmo. 2017. “The role of gender in the provision of public goods through tax compliance.” *Journal of Behavioral and Experimental Economics*, 71, 45-55.

Carrillo, Paul and M. Shahe Emran, 2018. “Loss Aversion, Transaction Costs, or Audit Trigger? Learning about Corporate Tax Compliance from a Policy Experiment with Withholding Regime.” MPRA Paper 87445, University Library of Munich, Germany.

Chang, Otto H. and Joseph J. Schultz Jr. 1990. “The income tax withholding phenomenon: Evidence from TCMP data.” *Journal of the American Taxation Association,* 12(1): 88-93.

Clotfelter, Charles T. 1983. “Tax evasion and tax rates: An analysis of individual returns.” *The Review of Economics and Statistics*, 65(3): 363-373.

Copeland, Phyllis V. and Andrew D. Cuccia. 2002. “Multiple determinants of framing referents in tax reporting and compliance.” *Organizational Behavior and Human Decision Processes*, 88 (1): 499-526.

Cummings, Ronald, Jorge Martinez-Vazquez, Michael McKee, and Benno Torgler. 2009. “Tax morale affects tax compliance: Evidence from surveys and an artefactual field experiment.” *Journal of Economic Behavior and Organization*, 70: 447-457.

Elster, Jon. 1989. “Social norms in economic theory.” *Journal of Economic Perspectives*, 3(4): 99-117.

Engström, Per, Katarina Nordblom, Henry Ohlsson, and Annika Persson. 2015. “Tax compliance and loss aversion.” *American Economic Journal: Economic Policy*, 7(4): 132-164.

Erard, Brian and Jonathan S. Feinstein. 1994. “The role of moral sentiments and audit perceptions in tax compliance.” Supplement to *Public Finance*, 49: 70-89.

Fehr, Ernst and Klaus Schmidt. 1999. “A theory of fairness, competition, and cooperation,” *Quarterly Journal of Economics*, 114(3): 817-868.

Feltham, Glenn D. and Suzanne M. Paquette. 2002. “The interrelationship between estimated tax payments and taxpayer compliance.” *Journal of the American Taxation Association*, 24(s-1): 27-45.

Fischbacher, Urs. 2007. “z-Tree: Zurich Toolbox for Ready-made Economic Experiments.” *Experimental Economics*, 10(2): 171-178.

Fochmann, Martin and Nadja Wolf. 2019. “Framing and salience effects in tax evasion decisions–An experiment on underreporting and over deducting.” *Journal of Economic Psychology*, 72: 260-277.

Gandhi, Ashvin and Michael Kuehlwein. 2014. "Reexamining Income Tax Overwithholding as a Response to Uncertainty." *Public Finance Review*, 44(2): 220-44.

Greiner, Ben. 2015. “Subject pool recruitment procedures: organizing experiments with ORSEE.” *Journal of the Economic Science Association*, 1: 114-125.

Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Valev. 2017. “The behavioralist as tax collector: using natural field experiments to enhance tax compliance.” *Journal of Public Economics*, 148: 14-31.

Harris, Louis and Associates, Inc., 1988. 1987 Taxpayer opinion survey. Conducted for the U.S. Internal Revenue Service, Internal Revenue Service Document 7292, Washington, DC, January 1988.

Hashimzade, Nigar, Gareth D. Myles, and Bin Tran-Nam. 2013. “Applications of behavioural economics to tax evasion.” *Journal of Economic Surveys*, 27(5): 941-977.

Holt, Charles and Susan Laury. 2002. “Risk aversion and incentive effects.” *American Economic Review*, 92 (5): 1644-1655.

Jones, Damon. 2012. “Inertia and over-withholding: Explaining the prevalence of income tax refunds.” *American Economic Journal: Economic Policy*, 4(1): 158-185.

Kahneman, Daniel and Amos Tversky. 1979. “Prospect Theory: An analysis of decision under risk.” *Econometrica*, 47: 263-291.

Kirchler, Erich and Boris Maciejovsky. 2001. “Tax compliance within the context of gain and loss situations, expected and current asset position, and profession.” *Journal of Economic Psychology*, 22: 173-194.

Kirchler, Erich and Ingrid Wahl. 2010. "Tax compliance inventory TAX-I: Designing an inventory for surveys of tax compliance." *Journal of Economic Psychology*, 31 (3): 331-346.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmual Saez. 2011. "Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark." *Econometrica*, 79(3): 651-692.

Martinez-Vazquez, Jorge, Gordon B. Harwood, and Ernest R. Larkins. 1992. "Withholding position and income tax compliance: Some experimental evidence." *Public Finance Review*, 20(2): 152-174.

McKee, Michael, Caleb A. Siladke, and Christian A. Vossler. 2018. "Behavioral dynamics of tax compliance when taxpayer assistance services are available." *International Tax and Public Finance*, 25(3): 722-756.

Pomeranz, Dina. 2015. "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax." *American Economic Review*, 105(8): 2539-2569.

Rees-Jones, Alex. 2018. "Quantifying Loss-Averse Tax Manipulation." *Review of Economic Studies*, 85: 1251-1278.

Robben, Henry S.J., Paul Webley, Henk Elffers, and Dick J. Hessing. 1990a. "Decision frames, opportunity and tax evasion." *Journal of Economic Behavior and Organization*, 14: 353-361.

Robben, Henry S.J., Paul Webley, Russell H. Weigel, Earl-Erik Wärneryd, Karyl A. Kinsey, Dick J. Hessing, Francisco Alvira Martin, Henk Elffers, Richard Wahlund, Luk Van Langenhove, Susan B. Long, and John T. Scholz. 1990b. "Decision frame and opportunity as determinants of tax cheating." *Journal of Economic Psychology*, 11: 341-364.

Saxonhouse, Gary. 1976. "Estimated parameters as dependent variables." *American Economic Review*, 66: 178-183.

Schadewald, Michael S. 1989. "Reference point effects in taxpayer decision making." *The Journal of the American Taxation Association*, 10(2): 68-84.

Schepanski, A. and T. Shearer. 1995. "A prospect theory account of the income tax withholding phenomenon." *Organizational Behavior and Human Decision Processes*, 63(2): 174-186.

Slemrod, Joel, Marsha Blumenthal and Charles Christian. 2001. "Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota." *Journal of Public Economics*, 79(3): 455-483.

Slemrod, Joel, Brett Collins, Jeffrey Hoopes, Daniel H. Reck, and Michael Sebastiani. 2017. "Does Credit-Card Information Reporting Improve Small-Business Tax Compliance?" *Journal of Public Economics*, 149: 1-19.

Torgler, Benno. 2002. "Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments." *Journal of Economic Surveys*, 16(5): 657-683.

Traxler, Christian. 2010. "Social norms and conditional cooperative taxpayers." *European Journal of Political Economy*, 26(1): 89–103.

U.S. Internal Revenue Service. 2016. "Tax Gap Estimates for Tax Years 2008 – 2010". https://www.irs.gov/pub/newsroom/tax%20gap%20estimates%20for%202008%20through%202010.pdf. Accessed May 2020.

U.S. Internal Revenue Service. n.d. "Pay as you go, so you won't owe: A guide to withholding, estimated taxes, and ways to avoid the estimated tax penalty. https://www.irs.gov/payments/pay-as-you-go-so-you-wont-owe-a-guide-to-withholding-estimated-taxes-and-ways-to-avoid-the-estimated-tax-penalty. Accessed May 2020.

Vossler, Christian A. and Scott M. Gilpatric. 2018. Endogenous audits, uncertainty, and taxpayer assistance services: Theory and experiments. *Journal of Public Economics,* 165: 217-229.

Vossler, Christian A. and Michael McKee. 2017. "Efficient tax reporting: The effects of tax liability information services." *Economic Inquiry*, 55(2): 920-940.

Vossler, Christian, Michael McKee, David Bruner and Michael Jones. "Taxpayer segmentation and tax reporting behavior: Effects of taxpayer services." Final Report on TIRNO-09-Z-00019 Task Order 003, May 2012.

Wooldridge, Jeffrey M. 2005. "Instrumental variables estimation with panel data." *Econometric Theory*, 21(4): 865-869.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). Cambridge, MA: The MIT Press.

Yaniv, Gideon. 1988. "Withholding and non-withheld tax evasion." *Journal of Public Economics*, 35(2): 183-204.

Yaniv, Gideon. 1999. "Tax compliance and advance tax payments: A prospect theory analysis." *National Tax Journal*, 52(4): 753.764.

Yitzhaki, Shlomo. 1974. "A note on income tax evasion: A theoretical analysis." *Journal of Public Economics*, 3 (2): 201-202.

Table 1: Experiment parameters

Parameter / variable	Value(s)
Income (expected value, EV)	Low: 1250 Medium: 1750 High: 2250 Uncertainty range: +/- 500 in withholding stage
Standard Deduction	250 or 500
Itemized Deduction (EV)	Low: 250 High: 500 Uncertainty range: +/- 250 in withholding stage
Audit Probability	10%, 30%, or 50%
Penalty Rate	300% on unpaid taxes
Tax Rate	50% on taxable income
Under-withholding Penalty	20% of amount owed at tax filing
Withholding Cost	10% of amount withheld
Tax Filing Time	120 seconds
Withholding Time	35 seconds
Liability information service	If available, cost is 50 lab dollars to acquire

Table 2: Treatment conditions

Treatment	Tax Withholding	Uncertain Income & Deductions	Liability Service Available	Public Good & Compliance Information	Number of Sessions
T1	Yes	Yes	No	No	3
T2	Yes	Yes	No	Yes	6
T3	Yes	Yes	Yes	No	3
T4	Yes	Yes	Yes	Yes	6

Table 3. Selected theoretical predictions: tax withholding and tax reporting

Income group	Tax withholding	Expected tax liability when filing	Predicted reported tax liability		
			10% audit rate	30% audit rate	50% audit rate
Low	0	437.5	0	0	400
Low	250	437.5	0	250	400
Low	500	437.5	0	312.5	437.5
Low	750	437.5	0	312.5	437.5
Low	1000	437.5	0	312.5	437.5
Middle	250	687.5	0	250	650
Middle	500	687.5	0	500	650
Middle	750	687.5	0	562.5	687.5
Middle	1000	687.5	0	562.5	687.5
Middle	1250	687.5	0	562.5	687.5
High	500	937.5	0	500	900
High	750	937.5	0	750	900
High	1000	937.5	0	812.5	937.5
High	1250	937.5	0	812.5	937.5
High	1500	937.5	0	812.5	937.5

Notes: Predicted reporting values are conditional on the withholding level and expected tax liability at the time of filing. These predictions assume: (1) the taxpayer optimally selects the allowable standard deduction; and (2) possible behavioral drivers (social norms, fairness, and reference dependence), which rely on unknown parameters, are absent. Shaded cells indicate optimal withholding and reporting values for the special case that expected tax liability is the same in the withholding and reporting stages. Additional details are provided in Appendix B.

Table 4. Variable descriptions

Variable Name	Description	Mean	S.D.
Tax Reported	Reported income minus reported deduction, multiplied by 50% tax rate	566.242	305.832
Liability Service	=1 if liability information service purchased	0.208	0.406
Tax Withheld	Amount of tax withheld	805.477	397.575
Expected Liability	Expected income minus expected deductions in reporting stage, multiplied by the tax rate of 50%	664.548	222.331
Tax Under-withheld	'Expected Liability' minus 'Tax Withheld', if >0; =0 otherwise	90.826	155.551
Tax Over-withheld	'Tax Withheld' minus 'Expected Liability', if >0; =0 otherwise	231.755	238.793
Under-withheld	=1 if 'Tax Under-withheld' >0	0.336	0.472
Over-withheld	=1 if 'Tax Over-withheld' >0	0.654	0.476
Compliance Norm	Lag of the compliance rate for income class multiplied by current round expected liability in reporting stage; =0 if 'Social Interactions'=0	376.749	340.736
Fairness	Lag of the mean taxes paid for the experiment group; =0 if 'Social Interactions'=0	376.327	285.459
Earned Income	Expected earned income in tax reporting stage	1776.901	432.290
High Standard	=1 if 500 Standard deduction allowed	0.504	0.500
Itemized	Expected allowable itemized deduction in tax reporting stage	365.013	142.455
Audit Rate	audit probability; .1, .3 or .5	0.299	0.164
Liability Service Not Purchased	=1 if available liability information service is available but not purchased	0.285	0.451
Social Interactions	=1 if info on compliance displayed & taxes partially reallocated	0.660	0.474
Tax Under-withheld (Info)	Same as 'Tax Under-withheld', but with expected liability measured pre-service and restricted to T3 and T4 participants	91.300	153.566
Tax Over-withheld (Info)	Same as 'Tax Over-withheld', but expected liability measured pre-service and restricted to T3 and T4 participants	216.319	232.588
Compliance Norm (Info)	Same as 'Compliance', but with expected liability measured pre-service and restricted to T3 and T4	388.226	332.077

	participants		
Earned Income (Info)	Same as 'Earned Income', but with expected liability measured pre-service and restricted to T3 and T4 participants	1779.167	416.778
Liability Service Available	=1 if liability information service available	0.492	0.500
Compliance Norm (WH)	Same as 'Compliance', but with expected liability measured using info in withholding stage	381.556	338.101
Earned Income (WH)	Expected earned income in withholding stage	1779.067	406.956
High Itemized	=1 if participant faced high itemized deduction range in withholding stage	0.495	0.500
Audited	=1 if selected for audit in prior round	0.299	0.458
Unexpected Tax Increase	'Expected Liability' minus Expected Liability (WH), if >0; =0 otherwise	12.263	29.360
Unexpected Tax Decrease	'Expected Liability (WH)' minus 'Expected Liability', if >0; =0 otherwise	14.294	32.420
Participant characteristics			
Risk Averse	=1 if selected sure bet in 70%, 80% or 90% lottery	0.389	0.487
Employed Full-time	=1 if participant employed full time	0.724	0.447
Female	=1 if participant is female	0.575	0.494
Age	Participant's age, in years	38.181	13.793
College Degree	=1 if participant has college degree	0.406	0.491
UT Lab	=1 if participant at UT experimental lab	0.533	0.499
Total Income	Participant's (individual) income, in $1000s	30.607	24.656
Non-matched Income	=1 if participant reported having non-matched sources of income	0.497	0.500
Asked for Advice	=1 if participant used tax advice from a non-tax professional for last tax return	0.221	0.415
Used Prep Service	=1 if participant used a professional tax preparation service to file last return	0.301	0.459
Reported Taxes Owed	=1 if participant reported additional taxes owed upon filing last tax return	0.126	0.332
Filed Jointly	=1 if married filing jointly on last tax return	0.265	0.442
Itemized Deductions	=1 if participant itemized on last tax return	0.243	0.429

Table 5. Tax reporting regressions

Dependent Variable: reported tax liability, in lab dollars (Tax Reported)			
	Model 1	**Model 2**	**Model 3**
Tax Withheld	0.18*** (0.02)		
Tax Under–withheld		–0.37*** (0.04)	–0.38*** (0.05)
Tax Over–withheld		0.06*** (0.02)	0.07*** (0.02)
Liability Service	101.00*** (19.35)	98.40*** (18.75)	140.82*** (38.43)
Liability Service × Under–withheld			68.42*** (22.05)
Compliance Norm	0.10*** (0.02)	0.10*** (0.02)	0.10*** (0.02)
Compliance Norm × Under–withheld			–0.02 (0.03)
Fairness	0.14*** (0.05)	0.14*** (0.05)	0.14*** (0.05)
Fairness × Under–withheld			0.01 (0.04)
Liability Service × Social Interactions			–103.63** (41.52)
Earned Income	0.29*** (0.01)	0.38*** (0.01)	0.38*** (0.01)
High Standard	–18.73*** (4.82)	–29.16*** (5.02)	–29.49*** (4.99)
Itemized	–0.10*** (0.02)	–0.14*** (0.02)	–0.14*** (0.02)
Audit Rate	210.68*** (24.82)	197.44*** (23.82)	196.04*** (23.57)
Liability Service Not Purchased	10.32 (19.05)	11.97 (18.67)	8.86 (18.86)
Social Interactions	–125.49*** (18.36)	–127.25*** (18.06)	–106.93*** (18.52)
Participant characteristics			
Risk Averse	34.67** (16.29)	35.96** (15.89)	37.40** (16.01)
Employed Full–time	–19.07 (22.64)	–16.19 (22.10)	–13.23 (22.22)
Female	41.80** (17.55)	42.86** (17.06)	47.67*** (17.21)
Age	3.04*** (0.98)	3.13*** (0.96)	2.90*** (0.97)
College Degree	–7.08 (21.27)	–7.34 (20.93)	–9.50 (21.13)
UT Lab	–63.27*** (18.95)	–60.86*** (18.36)	–57.65*** (18.47)
Total Income	0.18 (0.54)	0.06 (0.52)	0.08 (0.51)
Non–matched Income	–46.00** (18.39)	–46.51*** (17.92)	–44.81** (18.10)
Asked for Advice	44.27** (21.72)	41.98* (21.35)	37.72* (21.45)
Used Prep Service	31.99* (18.19)	32.67* (17.67)	30.37* (17.98)
Reported Taxes Owed	–37.85* (21.81)	–41.05* (21.24)	–37.93* (21.02)
Filed Jointly	39.63* (22.74)	38.06* (22.21)	41.83* (22.31)
Itemized Deductions	–51.21** (22.00)	–53.29** (21.33)	–52.34** (21.64)
Constant	–175.75*** (38.26)	–212.61*** (36.92)	–228.97*** (37.40)
Number of Observations	6022	6022	6022
R^2	0.715	0.719	0.722

*Notes: *, ** and *** denote estimates that are statistically different from zero at the 10%, 5% and 1% significance levels, respectively. Standard errors (parentheses) are clustered at the participant-level. Participant and decision round fixed effects are included in all models. Coefficients on variables that do not vary across rounds are estimated by regressing the estimated participant fixed effects on these variables.*

Table 6. Liability information acquisition regressions (Treatments 3 and 4 only)

Dependent Variable: =1 if liability service purchased; =0 otherwise (Liability Service)		
	Model 4	**Model 5**
Tax Withheld	0.009*** (0.003)	
Tax Under–withheld (Info)		–0.026*** (0.009)
Tax Over–withheld (Info)		–0.002 (0.004)
Compliance Norm (Info)	0.009* (0.005)	0.008* (0.005)
Fairness	0.009 (0.009)	0.000 (0.009)
Earned Income	–0.003 (0.003)	0.002 (0.003)
High Standard	–0.018 (0.012)	–0.021* (0.012)
Itemized	0.004 (0.005)	0.002 (0.005)
Audit Rate	0.299*** (0.057)	0.295*** (0.056)
Social Interactions	–0.082 (0.060)	–0.091 (0.059)
Participant characteristics		
Risk Averse	–0.005 (0.065)	–0.003 (0.065)
Employed Full–time	–0.040 (0.082)	–0.039 (0.082)
Female	–0.231*** (0.065)	–0.231*** (0.064)
Age	–0.002 (0.003)	–0.002 (0.003)
College Degree	0.139* (0.071)	0.140* (0.071)
UT Lab	–0.125* (0.070)	–0.120* (0.069)
Total Income	–0.000 (0.002)	–0.000 (0.002)
Non–matched Income	0.063 (0.063)	0.061 (0.063)
Asked for Advice	–0.010 (0.077)	–0.014 (0.076)
Used Prep Service	0.023 (0.070)	0.022 (0.070)
Reported Taxes Owed	–0.038 (0.090)	–0.044 (0.089)
Filed Jointly	–0.047 (0.085)	–0.050 (0.085)
Itemized Deductions	–0.008 (0.069)	–0.012 (0.068)
Constant	0.080 (0.117)	0.071 (0.118)
Number of Observations	2970	2970
R^2	0.634	0.633

*Notes: *, ** and *** denote estimates that are statistically different from zero at the 10%, 5% and 1% significance levels, respectively. Standard errors (parentheses) are clustered at the participant-level. Participant and decision round fixed effects are included in all models. Coefficients on variables that do not vary across rounds are estimated by regressing the estimated participant fixed effects on these variables. Variables measured in lab dollars are scaled by 100.*

Table 7. Tax withholding regression

Dependent Variable: Amount of tax withheld, in lab dollars (Tax Withheld)	
	Model 6
Liability Service Available	–8.95 (24.91)
Compliance Norm (WH)	0.02 (0.02)
Fairness	0.14** (0.06)
Earned Income (WH)	0.48*** (0.01)
High Itemized	8.33 (6.52)
Audit Rate	161.39*** (29.77)
Audited	15.01* (8.32)
Social Interactions	–25.28 (26.35)
Participant characteristics	
Risk Averse	68.12*** (25.15)
Employed Full–time	4.99 (31.57)
Female	98.89*** (26.21)
Age	5.17*** (1.43)
College Degree	–26.97 (29.26)
UT Lab	–19.41 (26.85)
Total Income	–0.89 (0.63)
Non–matched Income	–60.35** (25.87)
Asked for Advice	0.06 (32.55)
Used Prep Service	0.76 (28.63)
Reported Taxes Owed	2.75 (32.18)
Filed Jointly	–42.96 (32.40)
Itemized Deductions	3.61 (29.34)
Constant	309.52*** (55.16)
Number of Observations	6040
R^2	0.648

*Notes: *, ** and *** denote estimates that are statistically different from zero at the 10%, 5% and 1% significance levels, respectively. Standard errors (parentheses) are clustered at the participant-level. Participant and decision round fixed effects are included in all models. Coefficients on variables that do not vary across rounds are estimated by regressing the estimated participant fixed effects on these variables.*

Figure 1. Tax reporting and tax withholding