Tulane
University

Appeals to Social Norms and Taxpayer Compliance

James Alm
Tulane University
jalm@tulane.edu

William D. Schulze
Cornell University
wds3@cornell.edu
Jubo Yan
Cornell University
jy489@cornell.edu

Carrie Von Bose
Fors Marsh Group
cvonbose@forsmarshgroup.com

Abstract

We use laboratory experiments to examine the impact of appeals to social norms on the compliance decisions of individuals. We test the effects of two main types of social norms: "descriptive norms", or the type of behavior that is typical or most frequently enacted, and "injunctive norms", or the type of behavior that "constitutes morally approved and disapproved conduct". In addition, for injunctive norms we introduce approval-framed and disapprovalframed injunctive norm messages. Our results indicate that normative appeals generally have a modest and positive impact on tax compliance, if not always statistically significant. The magnitude of both approval- and disapproval-framed injunctive norm messages is an increase of around 2 percent in reported taxes.

Keywords: Tax compliance, social norms, experimental economics.
JEL codes: H2, H26, H3, C91.

APPEALS TO SOCIAL NORMS AND TAXPAYER COMPLIANCE

James Alm[a], William D. Schulze[b], Carrie von Bose[c], and Jubo Yan[d]

RUNNING HEAD: Appeals to Social Norms and Taxpayer Compliance

ABSTRACT: We use laboratory experiments to examine the impact of appeals to social norms on the compliance decisions of individuals. We test the effects of two main types of social norms: "descriptive norms", or the type of behavior that is typical or most frequently enacted, and "injunctive norms", or the type of behavior that "constitutes morally approved and disapproved conduct". In addition, for injunctive norms we introduce approval-framed and disapproval-framed injunctive norm messages. Our results indicate that normative appeals generally have a modest and positive impact on tax compliance, if not always statistically significant. The magnitude of both approval- and disapproval-framed injunctive norm messages is an increase of around 2 percent in reported taxes.

KEY WORDS: Tax compliance, social norms, experimental economics.

JEL CLASSIFICATIONS: H2, H26, H3, C91.

---

[a] Department of Economics, Tulane University, 6823 St. Charles Avenue, 208 Tilton Hall, New Orleans, LA, 70118. Phone: +1 504 862 8344. Email: jalm@tulane.edu.

[b] Charles H. Dyson School of Applied Economics and Management, Warren Hall, Cornell University, Ithaca, NY 14853. Phone: +1 607 227 9895. Email: wds3@cornell.edu.

[c] Fors Marsh Group, 1010 N. Glebe Road, Suite 510, Arlington, VA 22201. Phone: +1 571 858 3778. Email: cvonbose@forsmarshgroup.com.

[d] Charles H. Dyson School of Applied Economics and Management, Warren Hall, Cornell University, Ithaca, NY 14853. Phone: +1 607 227 9895. Email: jy489@cornell.edu.

1. INTRODUCTION

Improving tax compliance is a main goal of revenue agencies around the world. The standard policy tool has traditionally been increased enforcement efforts (e.g., larger penalties, higher audit rates). However, in recent years other policy tools have been suggested and tested. A recent field experiment carried out in the United Kingdom by Her Majesty's Revenue and Customs (HMRC) and HMRC's Behavioural Insights Team (BIT) finds that appeals to certain social norms in the form of letters sent to taxpayers requesting payment of unpaid taxes yields a significant positive response (Hallsworth, List, Metcalfe, and Vlaev 2017). A similar field experiment in Norway focused on individuals who owed taxes on foreign income, and tested several versions of a letter sent to taxpayers that encouraged, including some letters with moral appeals referencing the tax compliance behavior of the majority of citizens (Bott, Cappelen, Sorensen, and Tungodden 2017). These letters resulted in a small positive effect on the probability of reporting foreign income and a large positive effect on the amount of foreign income reported. A similar finding was made in a Washington state field experiment that also used normative appeals to promote increased reporting compliance by taxpayers subject to that state's Business and Occupation tax (i.e., an excise tax on gross revenues) (Iyer, Reckers, and Sanders 2010). However, evidence for social appeals is not clear-cut. An earlier field experiment conducted in Minnesota found no evidence that appeals to social norms had a significant impact on tax reporting compliance (Blumenthal, Christian, and Slemrod 2011), and other examples of field experiments report ambiguous effects of moral appeals, including Wenzel (2005, 2006), Wenzel and Taylor (2004), and Torgler (2004, 2012).[1]

In large part the mixed results on the impact of moral appeals on tax compliance reflect the fundamental difficulty of conducting empirical research on tax compliance: finding reliable

individual-level data. Finding hard and useful evidence on any tax compliance behavior is very difficult, for obvious reasons, and finding data that allow causal factors in the individual compliance decision to be examined is even more difficult. After all, tax evasion is illegal, and individuals have strong incentives to conceal any misreporting of tax liability, given financial and other penalties that are imposed on individuals who are found to be noncompliant on their taxes. For this basic reason, researchers have increasingly turned to controlled field experiments to test policy interventions, as with the studies noted earlier. These field experiments have a number of advantages.[2] The most obvious is their ability to identify causal factors in the compliance decision, given that they generate the randomization of variables necessary to create identification of the causal effects of policy interventions. In addition, they generate their own observable data on individual choices, they can be replicated, and they use "real people" making decisions in "real world" contexts. Even so, they seldom generate direct measures of evasion. They are also expensive. Of most importance, field experiments do not provide sufficient control over potentially relevant causal factors in the compliance decision, given the inherent and unavoidable inability to control the many unobserved and exogenous factors in the real world that can cause changes in behavior. These limitations suggest that both the internal validity (e.g., causality) and the external validity (e.g., generalizability) of field experiments are not always clear cut.[3] Indeed, these limitations likely go far in explaining the mixed results in field studies of appeals to social norms, along with of course differences in the specific design of these experiments.

As a result, we use in this paper laboratory experiments to test appeals to social norms as a means to improve tax payment compliance. Laboratory experiments seem particularly well-suited for this research. They allow researchers to establish control over all aspects of the

"microeconomic system" in which individuals make decisions (e.g., the institutions, the incentives), thereby allowing researchers to identify clearly the causal impact of policy changes on individual decisions by introducing these policy changes singly and exogenously in a controlled environment. They also generate direct measures of individual tax evasion, they are relatively inexpensive, and they can be easily replicated. However, laboratory experiments are sometimes viewed with suspicion. The most common criticism is that the student subjects typically used in experiments may not be representative of taxpayers. As a result, there is a concern that laboratory experimental results that rely upon student subjects cannot generalize to the population; that is, the external validity of laboratory experiments is sometimes questioned.[4] However, there is now much research that indicates subjects in tax compliance laboratory experiments exhibit behavior similar to "real" taxpayers (Alm, Bloomquist, and McKee, 2015).[5] To our knowledge, our paper is the first paper to use laboratory experiments to examine appeals to social norms as a strategy to improve tax compliance.

We formulate a model of the ways in which social norms affect an individual's compliance decision. We then test the predictions of this model using data from laboratory experiments in which different appeals to social norms are presented. Our experimental design follows the elements of much research on voluntary tax compliance (Alm, Jackson, and McKee 1992, 2009; Alm, McClelland, and Schulze 1992, 1997), incorporating additional features to improve parallelism with taxpayers' decision making in the naturally occurring world. Participants earn income by performing a task, they disclose income, and they face an audit process similar to that in the naturally occurring setting. The experimental instructions and the computer interface utilize tax language. The stakes are small, but the decision is simplified, implying that the ratio of decision costs and rewards parallels the naturally occurring setting

(Smith and Walker 1993). A key addition is that different social norm messages are sent to individuals in different treatments, to examine the impact of these messages on tax compliance. We test the effects of two main types of social norms: "descriptive norms", or the type of behavior that is typical or most frequently enacted, and "injunctive norms", or the type of behavior that "constitutes morally approved and disapproved conduct". In addition, for injunctive norms we introduce approval-framed and disapproval-framed injunctive norm messages.

Our results indicate that appeals have a modest and positive impact on tax compliance, if not always significant. The magnitude of both approval- and disapproval-framed injunctive norm messages is an increase of around 2 percent in reported taxes.

2. THEORETICAL BACKGROUND

The basic model of income tax compliance begins with the economics-of-crime model of Becker (1968), first applied to tax compliance by Allingham and Sandmo (1972). Here a rational individual i receives (or earns) income I_i and must choose how much of this income to report to the tax authorities R_i. The individual pays taxes at rate t on each dollar of reported income, and also incurs a cost C_i of preparing and filing a tax return. Unreported income is not taxed, but the individual may be audited with a fixed probability p, at which point the individual must pay unreported taxes plus a fine at rate f on all unreported taxes. The individual's income if caught underreporting is denoted I_C, and I_N is the individual's income if not caught underreporting. The individual weighs the benefits of successful underreporting against the risky prospect of detection and punishment in choosing the optimal amount of reported income R_i^*.

A standard approach is then to assume that the individual maximizes the expected utility of the evasion gamble. An alternative approach assumes that the individual maximizes the

expected value of the gamble, an approach that is implied by expected utility maximization when the individual is risk-neutral. It is this alternative approach that we explore in detail here.[6] Note that the assumption of risk-neutrality is not essential for our analysis. It serves mainly to simply the derivation of our theoretical results. It is also consistent with our experimental design as we do not elicit risk preferences from participants.[7]

More precisely, an individual i who chooses to report income R_i has expected value EV_i shown by equation (1):

$$EV_i = p\ I_C + (1\text{-}p)\ I_N, \tag{1}$$

where

$$I_C = I_i\text{ - }tR_i\text{ - }(1\text{+}f)\ t(I_i\text{ - }R_i)\text{ - }C_i$$

$$I_N = I_i\text{ - }tR_i\text{ - }C_i,$$

The expected value of the evasion gamble in equation (1) is the sum of two terms. The first denotes the individual's expected income if he/she is audited with probability p, in which case the individual pays taxes on reported income tR_i, incurs a cost of filing C_i, and is forced to pay all taxes on unreported income as well as a proportional fine f on unreported taxes resulting in income I_C. The second term denotes the individual's expected income if he/she is not audited with probability $(1\text{-}p)$ of this event, in which case the individual files a tax return at cost C, reports income R_i, and pays taxes tR_i, giving income I_N. Maximization of equation (1) by the choice of reported income indicates that individual i will optimally report all income if

$$p\ (1\text{+}f) > 1, \tag{2}$$

while the individual will report zero income if inequality (2) is reversed. The individual's decision is therefore all-or-none, reporting either all income or zero income. The presence of

risk-aversion modifies the all-or-none nature of the individual's decision, without changing the

basic comparative statics results.

This "portfolio" approach therefore gives the plausible result that compliance depends

positively upon enforcement, via the audit rate and the fine rate. Indeed, the central point of this

approach is that an individual pays taxes because – and *only* because – of this fear of detection

and punishment.[8]

Even so, this simple framework ignores many relevant considerations. Here we focus on

three aspects, all of which we incorporate later in the experimental design in order to improve

parallelism with taxpayers' decision making in the naturally occurring world.[9]

The <u>first</u> consideration is automatic reporting of employment income. A standard feature

of individual income tax systems is that a third party (e.g., the individual's employer or financial

institution) reports the relevant part of an individual's taxable income to the tax authority. The

automatic reporting of wage income to the tax authority reduces the individual's tax misreporting

opportunities. If an individual's income is automatically reported to the tax authority, it becomes

much more difficult to evade taxes on that portion of income, because any underreporting is very

likely to be detected. Under many tax regimes, some forms of income are subject to automatic

reporting, and others are not.

In addition to automatically reporting annual earnings to the tax authority, an individual's

employer also typically automatically withholds income taxes on wage income, which are then

paid directly to the tax authority. In many cases, the individual taxpayer chooses the rough level

of automatic withholding by completing a W-4 form, which declares a number of withholding

allowances; individuals who are self-employed are typically required to pay quarterly tax

payments based on their estimated annual income. The compliance effect of automatic

withholding, rather than being related to the probability of detection, is instead related to the framing of the tax decision and an individual taxpayer's ability to pay. If taxes are withheld throughout the year, the reference point becomes full payment of taxes, and any tax refund is seen as a benefit (rather than simply receiving money that one already earned). However, any tax payment still due may remain unpaid if a taxpayer has not appropriately saved for this eventuality. Additionally, taxpayers may be less willing to pay additional taxes owed because the payment is more salient at the time of filing than it is throughout the year as taxes are automatically withheld.

Relatedly, the individual may face some uncertainty about the appropriate level of withholding to choose, due either to uncertainty about his or her income or uncertainty about the true tax liability. An individual's income may vary from year to year. Furthermore, the tax code is undeniably complex, and the eligibility for allowable deductions and credits is frequently open to interpretation. The effects of such uncertainty can lead to either increasing or decreasing compliance.[10]

To address this first consideration, we partition reported income R_i between what the individual chooses to have the employer withhold automatically (R_{iEW}) and what the individual self-reports (R_{iSR}), where $R_i = R_{iEW} + R_{iSR}$. The individual now chooses both the amount of income subject to employer withholding and the amount of self-reported income when maximizing expected value. Note that the expected value still depends simply upon the choice of reported income R_i, given that the same tax and penalty rates apply equally to R_{iEW} and to R_{iSR}.

The second consideration is the uses to which tax payments may be put. The standard portfolio model of individual behavior ignores these uses. However, taxes are used to finance a range of government services, especially public goods like national defense, parks, and

infrastructure. Economists have long argued that the private provision of a public good will be inefficiently low because each individual will have an incentive to free ride on the private purchases of others. Because a public good is nonexclusive, an individual who does not contribute to its provision cannot be excluded from its use. Also, because all individuals face this same incentive, individuals will not voluntarily contribute to a public good. Put differently, public goods create a *social contribution dilemma* in which individual gain runs contrary to the collective good (Dawes 1980). However, there is also much work that argues that voluntary provision of a public good may not always play as a prisoners' dilemma game (Ledyard 1995). This possibility arises especially when an individual's optimal decision depends upon the actions that he or she expects others to follow. Under some circumstances, full voluntary contributions (or cooperation) may be the dominant individual strategy. This work therefore suggests that individuals pay taxes (voluntarily) because they value the goods provided by government and because they recognize that their payment may be necessary to get others to contribute (Alm, McClelland, and Schulze 1992).

The presence of a public good that is financed from individuals' taxes can be easily incorporated in the standard portfolio model. All that is required is the addition to both I_C and I_N of a term that represents the value of the public good financed by total group taxes and returned to each individual. For each individual i, this term equals $[mst(\sum_{j \neq i} R_j + R_i)]$, where total group taxes $t(\sum_{j \neq i} R_j + R_i)$ equal the tax rate t multiplied by the amount of income reported by individual i (R_i) plus the amount reported by all other j members of the group other than individual i ($\sum_{j \neq i} R_j$), m is a *group surplus multiplier* applied to the group's total taxes to capture the consumers' surplus derived from a public good, and s is the share of individual i in the resulting public good.

It is readily shown that individual i is more likely to pay taxes with higher values of m, higher values of s, and higher levels of contributions by other group members $(\sum_{j \neq i} R_j)$.

The <u>third</u> consideration is the role that social norms play in individuals' reporting and compliance decisions (Alm, McClelland, and Schulze 1999). The portfolio model focuses entirely on the self-interested financial considerations of reporting. However, there is abundant evidence that individuals are motivated not simply by self-interest but also by group notions like social capital, social customs, social networks, fairness, trust, reciprocity, intrinsic motivation, or tax morale, and by individual notions of guilt, shame, morality, or altruism. In short, individuals are not always the rational, outcome-oriented, self-controlled, selfish, and egoistic consumers envisioned by much of our standard theory.

In particular and building on work in the psychology of taxation (Lewis 1982; Kirchler 2007), there is much evidence of what may be termed a *social norm* of tax compliance (Elster, 1989). Social norms have been studied extensively in the psychology literature (e.g., Cialdini, Reno, and Kallgren 1990; Schultz et al. 2007), and it has been found that individuals tend to feel discomfort when they perceive that their attitudes or behaviors differ from the prevailing norms of their cohort. A social norm therefore represents a pattern of behavior sustained in part by social approval or disapproval, either through internalized feelings of guilt or shame for not complying with norms or through external sanctions or punishments administered by the group. The concept of a social norm is related to and consistent with a range of other individual emotional and motivational factors that influence tax compliance behavior.[11] An individual has an incentive to comply as long as he or she believes that compliance is the social norm. Conversely, if noncompliance becomes pervasive, then the social norm of compliance disappears, and individuals do not suffer the negative psychological effects of acting contrary to

the norm. These effects of social norms suggest that knowledge about other taxpayers' decisions affects one's own compliance decision.

There are many ways in which behavioral notions relating to a social norm can be introduced. Perhaps the simplest way is suggested by the work of Kahneman and Tversky (1979), who incorporate what they term a *reference point* in their prospect theory by assuming that an individual suffers a loss in utility if he or she does not achieve some given level of income or wealth defined by the reference point. The reference point is traditionally taken to be the current asset position[12], but it also can be created through social comparison. If we take the social norm as a reference point, individuals perceive losses relative to this point to be worse than equivalent gains. For a social norm of full tax compliance, individuals will maximize utility relative to the reference point by reporting all income and paying all taxes. An individual who declares less than full income and pays less than full taxes has higher net income on average but faces the risk of an unfavorable audit. Because the reference point is full tax compliance, the penalties produced by an audit create a much larger loss in utility than the small expected gains produced through tax evasion, since losses are weighted more heavily than gains. Prospect theory suggests that a reference point created by social norms is one way to explain why knowledge of social norms should be expected to increase tax compliance.

More formally, assume that each individual i suffers a psychological loss proportional to undisclosed taxes, equal to $[\gamma_i t(I_i - R_i)]$ in expected income , where the coefficient γ_i measures as a fraction of income how much individual i would pay to avoid the loss associated with each dollar of unreported taxes. It is straightforward to demonstrate that an individual is more likely to report more income in the presence of this psychological loss and that reported income increases with an increase in γ_i. Clearly, γ_i is likely to be sensitive to the social norm of tax compliance. The stronger

the social norm, the more deviant the behavior of a noncompliant individual becomes, and the more loss the individual feels.

Putting these three considerations together, the individual's calculus is more complicated. The individual now chooses the amount of income to withhold and the amount of income (out of that which is not subject to automatic reporting) to report to the tax authority in order to maximize expected value, taking into account the existence of a public good financed by all individuals' reported income and the social norm of compliance.[13] The condition for reporting all income is now modified from inequality (2) to

$$p\,(1+f) + ms > 1 - \gamma_i\,. \tag{3}$$

The individual's decision remains all-or-none. Even so, it is clear that inequality (3) is more easily satisfied than inequality (2), so that it is more likely that the individual will report all income.

Indeed, in the face of these many elements, it is straightforward to show that an individual will increase reported income if there is an increase in the enforcement variables (e.g., the audit probability and the penalty rate).[14] The individual will also increase reported income if there is an increase in the payoff from the public good. Finally, an appeal to a social norm will also increase the likelihood that the individual reports all income because an appeal to a social norm will increase the psychological cost of noncompliance.

This framework suggests several testable hypotheses[15]:

Hypothesis 1 (Enforcement): An individual's reported taxes will increase with greater enforcement (e.g., a higher penalty rate, a higher audit rate).[16]

Hypothesis 2 (Payoffs): An individual's reported taxes will increase with greater group payoffs (e.g., a larger public good).

Hypothesis 3 (Appeals to Social Norms – Reported Taxes): An individual's reported taxes will increase with appeals to social norms.

Hypothesis 4 (Appeals to Social Norms – Withholding): An individual's withheld taxes will increase with appeals to social norms.

The next section presents our experimental design for examining these impacts, especially the impacts of appeals to social norms via messages.

3. EXPERIMENTAL DESIGN

General Experimental Features

We utilize a laboratory experiment to examine the impact of social norms on tax withholding and reporting decisions. In this experiment, social norms are induced in three treatments in order to observe resulting differences in tax decisions. The experimental design captures the essential features of the voluntary income reporting and tax assessment system used in many countries. Human participants in a controlled laboratory environment perform a task that pays them income, and they also receive a random income component. The participants must decide how much of their total income will be subject to automatic income tax withholding, and then they must decide how much of their total income to report to a tax agency. Taxes are paid on reported income only. Any unreported income may be discovered via a random audit and then the individual must pay the owed taxes plus a fine based on the unpaid taxes. The probability of detection is fixed and known to the individual and is independent of the individual's decisions. Subjects are fully and accurately informed about the various features of the experimental setting (e.g., tax rates, penalty rates, audit rates, public good payoffs, and the like). This withholding, reporting, audit, and penalty process is repeated over a number of rounds, each representing a tax period. At the completion of the experiment, all participants are paid in cash an amount based upon their laboratory market earnings, converted to U.S. dollars.

Participants are recruited from a pool of undergraduate students at two major universities, one public (Appalachian State University) and one private (Cornell University), using the Online Recruiting System for Experimental Economics (ORSEE) developed by Greiner (2004) and Sona System software, respectively. Upon arrival at the laboratory, participants are assigned to a computer station, which assures privacy; each session had 20 participants, with the exception of one session for Treatment 2 at Appalachian State University that had only 19 participants. Basic instructions are provided via a hard copy and also via a series of screen images (see Appendix A for representative instructions). The instructions use tax language, rather than more "neutral" terms.[17] After reading the instructions, participants are allowed to ask questions. Decisions are made privately, and participants are not allowed to communicate with other group members during the session. Participants are informed (via the consent sheet) that all responses are anonymous and that no individual identification will be collected. Because payoffs depend on multiple decisions made by each participant, as well as on the random audit process, the person making the payments cannot assess participant decisions from the level of the aggregate payoffs at the end of the session.

Participants are not told the exact duration of the experimental session, which is predetermined to last for 20 real rounds. Including instructions, practice rounds, and the real rounds, sessions take on average 75 minutes to complete. Participant earnings range from $26 to $32, depending upon subject performance.

Once subjects complete the informed consent and are taken through the detailed instructions, they follow five steps, which are briefly described as follows:

Step 1. Each participant performs a task for which they are paid to provide taxable earnings on each round. The task involves estimating the number of gumballs (or marbles) in a

jar at the front of the lab. The decision screen associated with the guessing task is shown in Figure A1 in Appendix A. An exact estimate results in the maximum earnings (10,000 lab dollars). Subjects' earnings are reduced from this maximum by a linear function (common knowledge) of 50 lab dollars for each gumball or marble their estimate varies from the actual number in the jar, but participants are guaranteed a minimum fixed income of 5,000 lab dollars in each round. This amount represents the fixed component of their income, which is in effect for the duration of the experiment. As most people who earn self-reported income are likely to face some variation in year over year income, the subject also receives a random income component in each round, which is added to the fixed income from the earnings task. The random portion of the income follows a uniform distribution that has a lower bound of 5,000 lab dollars and an upper bound of 10,000 lab dollars.

Step 2. After completing the earnings task, the tax withholding and reporting rounds proceed. Each round represents a tax year. The subjects make their withholding decision while only knowing their fixed income for the year; the variable portion of income is revealed on the next decision screen. There is a penalty for under-withholding as applied by the tax agency, and an opportunity cost (foregone consumption or interest) for over-withholding. Consistent with experimental methods, these values are induced by the experimental setting. Participants choose their tax withholding amount from a menu for that round. In the first ten rounds of the experiment, no social norm messages are presented. In the second ten rounds, social norm messages are presented prior to choosing the withholding rate for individuals in one of the three treatment groups. The control group does not receive social norm messages at any point during the experiment. The order of receiving messages is not reversed because such messages are normally introduced in the field after a period where no such message has been provided.

Step 3. The tax reporting phase of the tax period requires the subjects to claim a deduction to determine their tax liabilities. Taxable income is reported income minus deduction. The tax form is filed. Collected taxes fund a public good (implemented as a transfer payment to all participants). Figures A2 and A3 in Appendix A show the withholding and tax filing screens.

Step 4. After the results have been examined for the round, the computer randomly selects individuals for audit. This random process is independent across subjects, and the probability of being selected is common knowledge.[18] Audits work perfectly (i.e., all unpaid taxes are detected), and a penalty plus any unpaid taxes are collected from the subject. Under-withholding is also penalized at this point. All subjects receive a public good benefit that is calculated by multiplying the total amount of taxes paid by 1.4 and dividing equally among all subjects.[19] The subjects then get a final summary screen that shows their earnings (including penalty costs and transfer payments) for the round. Figure A4 in Appendix A shows the outcome screen.

Step 5. At the conclusion of the study (20 paid rounds), participants receive their final balance in cash and complete a short, 5-minute survey that asked questions regarding their experience in the experiment, demographic information, and variables relating to altruistic attitudes and behaviors. These variables have been linked to tax compliance in past research (Dwenger et al. 2016; Alm and Christian, 2014). The survey questions are in Appendix B.

Experimental Economics Facilities

We use two laboratories that meet the relevant requirements for technology, accessibility, and participant diversity. Both locations use the same software, and the physical layouts of the labs are parallel.

Cornell University. The Laboratory for Experimental Economics & Decision Research (LEEDR) was established in 1996 in the Dyson School of Applied Economics & Management at Cornell University. It was the first experimental economics laboratory among Ivy League universities. The facility is dedicated to research, and it has 24 subject computers equipped with privacy shields, two-monitor computers, a dedicated server in the lab, and the latest audio and visual equipment. The lab maintains a list of several thousand potential student participants drawn from departments across the Cornell campus in Ithaca, New York.

Appalachian State University. The experimental lab at Appalachian State (AppEEL) has been used for previous IRS-funded research (including joint work with the University of Tennessee and Appalachian State University). The AppEEL facility has 25 participant stations. The lab uses computerized participant recruiting and has a database of over 1,000 potential student participants.

Participants

Two independent experiments were conducted at Appalachian State University and Cornell University. Subjects were drawn from their respective student databases. Because the objective of the study is to test the influence of messages reminding participants of social norms, it should be noted that business and economics students show different levels of other-regarding behavior than students with majors such as sociology, music, and art and that preferences also can vary depending on political views or associations.[20] The LEEDR and AppEEL databases incorporate information for potential participants on these attributes as well as on gender, age, and the like, which allows us to draw samples that are stratified to better determine the influence of personal characteristics on response to reminders of social norms relating to tax compliance.

At the time of this study the AppEEL subject pool was just over 1000 active student subjects. Female subjects were 58 percent of the pool, and 18 percent of subjects in the pool listed their major field of study as business administration, with psychology (at 10 percent) as the next largest field. The remaining pool was fairly evenly distributed over a variety of arts and science majors, education majors, and a small subset from music and fine arts majors. The sampling at both sites used a broad range of undergraduate student subjects. See Table 1.

Experimental Treatments: Social Norm Messages

Our objective is to examine the effects of social norms on individual reporting decisions. To establish a baseline, we conduct laboratory experiments using student subjects in which no normative messages are given to subjects. We then introduce three different social norm messages that vary in the type of normative information communicated. For each of the four treatments, we recruited 120 participants, including the control treatment. A total of 479 participants completed the experiment in a total of 24 experiment sessions, split equally between the two sites at Cornell University and Appalachian State University.[21]

Testing a large number of potential statements using only tax compliance economics experiments to evaluate them would be very costly, so we used focus groups to screen alternative messages (using hypothetical responses) that might be used in the economics experiments to follow. The focus groups consisted of undergraduates at Cornell University and undergraduates at Appalachian State University. A recent summary of the literature on social norms in Schulze and Hoffer (2014) forms one basis for developing such messages in addition to results from existing field experiments.

Research on the psychology of social norms suggests that the impact of social norm messages can differ depending on the type of norm targeted. In particular, two types of norms appear to drive behavior: "descriptive norms", or the type of behavior that is typical or most frequently enacted, and "injunctive norms", or the type of behavior that "constitutes morally approved and disapproved conduct" (Cialdini et al. 1990). Although the two types of norms are distinct constructs (Park and Smith 2007), their isolated use may yield different results depending on contextual information. In particular, devoid of an injunctive norm, descriptive norms that describe the average behavior of others can have a "boomerang effect". Consider a social norm message that describes the frequency of an undesirable behavior, such as underage alcohol consumption on a college campus. In this example, let that frequency be 35 percent. For many students on campus who are aged 18 to 20, this percentage might be *lower* than their existing perception; for others, this percentage could actually be *higher* than their baseline perception. In the latter case, a message that focuses solely on the descriptive norm could result in increased alcohol consumption among some members of the target population (i.e., among underage students at the university).

However, when a message couples a descriptive norm with an injunctive norm, this type of undesired effect can be avoided. Indeed, there is evidence that the potential boomerang effect of descriptive norms can be eliminated by the addition of information about the level of social approval or disapproval of a behavior (Cialdini et al. 1990; Schultz et al. 2007). For instance, if the frequency from the previous example was qualified by a message noting that most students at the university disapprove of underage alcohol consumption, the potential for the undesired effect is likely to be substantially mitigated.

This research suggests that those who violate the norm of tax compliance are less likely to violate the norm if an appeal is made, but those who are in compliance may actually be more likely to violate the norm, unless their behavior is rewarded with some indication of social approval (Schultz et al. 2007; Irwin and Simpson 2013). Additionally, Schulze and Wansink (2012) show that positive emotions can offset the impact of negative emotions on decisions[22], suggesting that messages with approval framing are likely to be effective even if tax compliance invokes some negative emotions.

The social norm messages resulting from these focus groups and used in the experiments are as follows:[23]

Treatment	Group name	Test phrase
1	Control	Note: No information is presented.
2	Descriptive Norm	"In a previous session of this experiment, a large majority of Appalachian State University/Cornell University students withheld enough earnings to pay their entire tax liability, and 12% did not. This is very similar to the country as a whole where 3 in 4 Americans withhold enough taxes throughout the year to pay their entire tax liability, and 1 in 4 does not."[24]
3	Injunctive Norm, Approval-framed	"In a previous session of this experiment, a large majority of Appalachian State University/Cornell University students withheld enough earnings to pay their entire tax liability. This is very similar to the country as a whole where 3 in 4 Americans withhold enough taxes throughout the year to pay their entire tax liability. 90% of Americans say that personal integrity is a big reason why they comply with tax regulations,[25] and those who withhold enough taxes have a 97% tax compliance rate."[26]
4	Injunctive Norm, Disapproval-framed	"In a previous session of this experiment, a minority of Appalachian State University/Cornell University students did not withhold sufficient funds to pay their entire tax liability. This is very similar to the country as a whole where only 1 in 4 Americans still owe taxes at the time of filing. 88% of Americans agree that any type of tax cheating is unacceptable,[27] and people who do not withhold enough earnings to pay all of their taxes are 4 times more likely to cheat on their taxes."[28]

Using these social norm messages, our basic research question is simple: "Do social norm messages increase compliance in a lab experiment setting?"

4. RESULTS

We first present simple descriptive statistics from the various sessions (Table 2). For both

Tax Paid (or the amount of reported taxes by the subject) and *Withholding Amount* (or the amount of individual tax withheld), we report the average level of the variable, averaged across all subjects and all (relevant) rounds, along with its standard error. The *t*-tests show that subjects reported significantly less taxes in the second half than in the first half in all treatments. In terms of the withholding amount, the injunctive norms significantly affect subjects' decisions, with the different framings driving the different withholding amounts.

Note that the dominant strategy equilibrium for a risk-neutral individual who maximizes the expected value of the compliance gamble is noncompliance, given the levels of tax rates, audit rates, and fine rates. In fact, we observe substantial levels of noncompliance. Even so, we also observe many individuals who comply fully. Put differently, individual behavior is often consistent with all-or-none behavior, as suggested by our theoretical framework.

Figures 1 and 2 show mean withholding and mean taxes paid by round in the control treatment for Appalachian State participants (orange squares) and Cornell participants (green circles), respectively. Note that there is an approximately linear decay across rounds in the case of taxes paid, suggesting that the round should be included as an explanatory variable to control for learning.

We use fixed effects regressions to control for repeated observations from the same participant, in which we estimate the amount of taxes paid (*Reported Taxes*) by each participant in each round against the following variables, as suggested by our theoretical framework:

- *Tax Obligation (*the "true" amount owed)
- *Round*
- *Treatment 2 Dummy Variable (Descriptive Norm)*
- *Treatment 3 Dummy Variable (Injunctive Norm, Approval-framed)*
- *Treatment 4 Dummy Variable (Injunctive Norm, Disapproval-framed)*
- *Penalty* (paid in the previous round)
- *Public Good Benefit* (paid on the previous round).

Of most interest, we include the various treatment variables as dummy variables that measure the treatment effects of the different social norm messages. *Treatment 2 Descriptive Norm* measures the reporting impact of a simple descriptive norm on amount of reported taxes, *Treatment 3 Injunctive Norm – Approval Frame* measures the impact of an approval-framed injunctive norm, and *Treatment 4 Injunctive Norm – Disapproval Frame* measures the impact of a disapproval-framed injunctive norm. In all cases, the treatment impact is measured relative to experiments with no social norm message of any type as well as the first half of the rounds in the respective treatment. We include a control for the amount of any audit penalty assessed in the previous round to account for any temporary increase in compliance brought on by the increased salience of being audited, and a control for the amount of the public good benefit received in the previous round to account for the benefits associated with paying one's taxes. We also include a variable for the individual's "true" tax liability (*Tax Obligation*, or the amount of taxes that should be paid on the individual's full, or "true" income), and a variable for the round of the experiment to capture any possible change in compliance as the individual gains experience or learning as the experiment proceeds (*Round*). We estimate the models using panel methods by including subject fixed effects, with standard errors clustered at the subject level.

These estimation results are presented in columns (1) and (2) of Table 3. As expected, *Tax Obligation* has a positive and statistically significant impact on *Reported Taxes*. Its coefficient of around 0.9 in both specifications suggests a tendency to underpay true taxes. *Round* exhibits a negative trend, suggesting that participants are learning to cheat across rounds of the experiment. Surprisingly, *Penalty* has a marginally significant but negative correlation with the amount of *Reported Taxes* paid in the current round. This result may indicate that noncompliance is "sticky"; that is, if a taxpayer fails to report income in one round, he or she is

likely to fail to report income in subsequent rounds, even immediately after being assessed an audit penalty (Hypothesis 1). This result may also indicate that subjects exhibit some form of "gambler's fallacy" (or "bomb-crater effect"), in which they anticipate that they will not be audited again even after an initial audit.[29] The *Public Good Benefit* paid in the previous round is positive but weakly significant, so that individuals pay more in taxes when they receive more for their tax payments (Hypothesis 2).[30]

Of most importance are the treatment variables. The coefficient estimate for *Treatment 2 Descriptive Norm* is positive but not significant. The coefficient estimates for each of the two injunctive norm variables are positive, large, and statistically significant (*Treatment 3 Injunctive Norm, Approval-framed* and *Treatment 4 Injunctive Norm, Disapproval-framed*), indicating that these appeals to social norms have a positive impact on *Reported Taxes* (Hypothesis 3).

The main conclusions are that all three social norm variables have a positive impact on compliance, with both the approval- and the disapproval-framed injunctive social norm messages statistically significant in increasing compliance. Note also that the effects of the approval- and disapproval-framed messages are statistically identical.[31]

We also report in Table 3 several different specifications to examine the robustness of our main results.[32] In column (3) we include the withholding amount and a dummy variable that indicates whether the withholding is "enough" (or sufficient) to cover the tax liability. The results show that withholding enough significantly increases *Reported Taxes* and also that the withholding amount does not affect the amount of reported taxes; the results in column (3) also show that the treatment dummy variables for both injunctive norms remain significant at the 1 percent level, suggesting that the social norm messages influence tax payments through changes in both the withholding behavior and the final payment behavior. Further, in column (4) of Table

3 we focus on those who cheated for at least 5 rounds in the first 10 rounds of the experiment. As expected, the effects of the social norm messages are larger on this subgroup of subjects. Even so, only the two injunctive norm treatment messages are statistically significant for this subgroup of subjects.

In additional robustness tests, we find similar results when we replace the dependent variable of Table 3 (or *Reported Taxes* with the amount of evaded income (or *Unreported Income*). See Table 4. Again, the two injunctive norm treatment messages have a negative and statistically significant impact on evaded income, as expected. The descriptive norm treatment variable also has a negative impact on evaded income, but its coefficient is not significant.

In Table 5 we examine the effects of various demographic variables on the determinants of paid tax: gender, race, religion, and volunteering behavior, reporting only the effects of the treatment variables. In most all cases, the treatment variables have positive and significant impacts on Reported Taxes, with effects that vary in size by demographic characteristics. *Females* are more likely to be influenced by appeals to social norms than Males. The impacts differ somewhat by *Race* (e.g., dummy variables for *White* versus *Non-white*), with somewhat larger impacts of injunctive norms for *Non-white* versus *White* subjects. The impacts also differ somewhat in size by *Religion* (e.g., dummy variables for *Christian* versus *Other* religions). Subjects who tend to engage in more *Volunteering Behavior* (*Below Median* versus *Above Median*) seem somewhat more responsive to injunctive social norm appeals.

We also estimate the determinants of *Withholding Amount* choices using the same basic specifications (Table 6), where *Withholding Amount* measures the individual's choice of the amount of taxes to withhold (Hypothesis 4). The only significant variable in the two specifications is *Constant*. As for the social norm variables, the approval-framed message

(Treatment 3) increases withholding, although the effects are not significant. Surprisingly, the disapproval-framed message (Treatment 4) decreases withholding, as does the descriptive norm (Treatment 2). Even so, as noted above, all three messages still have a positive impact on compliance as measured by *Reported Taxes*. Since increased withholding is likely a desirable outcome for the tax authority, approval-framed messages seem to be more desirable, particularly for scenarios in which taxpayers make active decisions about tax withholding (such as self-employed individuals in the U.S. who make quarterly estimated tax payments).

These results suggest that the social norm messages are effective mainly because of the injunctive component and that messages containing the descriptive element alone may not lead to significant changes in behavior. More research is needed to examine the seemingly conflicting effects of the disapproval-framed injunctive norm message. This could be evidence of a "boomerang effect", in which participants initially believed that a higher proportion of other participants withheld enough taxes. It is also possible that the approval-framed messages more clearly link withholding behavior to compliance behavior and thus are more effective at improving both behaviors. In addition, the disapproval-framed message uses the word "cheating", which participants may have linked with the tax payment only, whereas the approval-framed message uses the word "compliance", which participants may have associated with the entire tax filing process. Furthermore, the injunctive portion of each message was only focused on tax compliance since it is difficult to imagine that the majority of Americans have strong feelings about tax withholding. A potential policy implication is that it may be possible to encourage a desired behavior (e.g., sufficient tax withholding) by referencing a related social norm (e.g., tax compliance), but messaging should be carefully crafted and tested. Future experiments could include messages that address compliance only, compliance plus withholding,

and withholding only.

5. CONCLUSIONS

Our experiments show a small but significant impact of social norm messages on tax compliance. The magnitude of both approval- and disapproval-framed injunctive norm messages in these experiments is an increase of around 2 percent in taxes paid. Of course, quantitative extrapolation directly from laboratory experiments to actual tax policy is somewhat risky. Even so, if a similar response were to occur for the U.S. tax system with roughly a trillion dollars in taxes collected, the result would be an increase in tax revenue of around $20 billion.

We believe that our laboratory results make several significant contributions to the existing work on appeals to social norms, most of which is based on field studies. First, and most importantly, our laboratory experiments are able to examine the impacts of social norm appeals in an environment that closely parallels the real-world environment in which individuals make their compliance decisions, and to examine the effects of these appeals in an environment that fully controls for other potential impacts on compliance decisions. Second, our results suggest that additional laboratory – and field – tests of social norm messages may be worthwhile, especially given the contradictory and inconsistent results of existing field studies. Third, an implication of our research is that, for taxpayers who prepare their own taxes, placement of social norm message on the tax forms themselves, as done in this experiment, may be an inexpensive way of reaching taxpayers in comparison to mailings or other forms of communication such as advertising. Fourth, an important feature of our experimental design is that the social norm messages were not a one-time occurrence (as with existing field studies of tax compliance), but were repeated round after round. This feature may help to mitigate non-cooperative behaviors that might otherwise emerge in a social dilemma setting. It is also a

feature that is largely unexamined in existing field studies, at least in a tax compliance setting.[33]

Finally, our results indicate that the "boomerang effect" appears to be a serious concern for social norm messages, as illustrated by the effect on withholding of the disapproval-framed message. All of these issues require further study.

REFERENCES

Allcott Hunt. 2011. Social norms and energy conservation. *Journal of Public Economics* 95 (9–10): 1082-1095.

Allingham, Michael, and Agnar Sandmo. 1972. Income tax evasion: A theoretical analysis. *Journal of Public Economics* 1 (3-4): 323-338.

Alm, James. 1988. Uncertain tax policies, individual behavior, and welfare. *The American Economic Review* 78 (1): 237-245.

Alm, James. 2012. Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field Studies. *International Tax and Public Finance* 19 (1): 54-77.

Alm, James. 2019. What motivates tax compliance? *Journal of Economic Surveys* 33 (2): 353-388.

Alm, James, Kim Bloomquist, and Michael McKee. 2015. On the external validity of tax compliance experiments. 2015. *Economic Inquiry* 53 (2): 1170-1186.

Alm, James, Todd Cherry, Michael Jones, and Michael McKee. 2010. Taxpayer information assistance services and tax reporting behavior. *Journal of Economic Psychology* 31 (4): 577-586.

Alm, James, Todd Cherry, Michael Jones, and Michael McKee. 2012. Social programs as positive inducements for tax participation. *Journal of Economic Behavior & Organization* 84 (1): 85-96.

Christian, Roberta Calvet, and James Alm. 2014. Empathy, sympathy, and tax compliance. *Journal of Economic Psychology* 40: 62-82.

Alm, James, Jeremy Clark, and Kara Leibel. 2016. Socio-economic diversity, social capital, and tax filing compliance in the United States. *Southern Economic Journal* 82 (3): 725-747.

Alm, James, Betty R. Jackson, and Michael McKee. 1992. Institutional uncertainty and taxpayer compliance. *The American Economic Review* 82 (4): 1018-1026.

Alm, James, Betty R. Jackson, and Michael McKee. 1993. Fiscal exchange, collective decision institutions, and tax compliance. *Journal of Economic Behavior & Organization* 22 (4): 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. 1992. Why do people pay taxes? *Journal of Public Economics* 48 (1): 21-38.

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

Andreoni, James, Brian Erard, and Jonathan Feinstein. 1998. Tax compliance. *The Journal of Economic Literature* 36 (2): 818-860.

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

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.

Bott, Kristina M., Alexander W. Cappelen, Erik Ø. Sørensen, and Bertil Tungodden. 2017. You've got mail: A randomised field experiment on tax evasion. Norwegian School of Economics Department of Economics Discussion Paper 10/2017. Bergen, Norway.

Braithwaite, Valerie. 2009. *Defiance in Taxation and Governance – Resisting and Dismissing Authority in a Democracy*. Cheltenham, UK and Northhampton, MA: Edward Elgar Publishing.

Bühren, Christoph, and Thorben Kundt. 2013. Worker or shirker – Who evades more taxes? A real effort experiment. MAGKS Joint Discussion Paper Series in Economics 26-2013. Marburg, Germany: Philipps-Universität Marburg.

Camerer, Colin F. 2015. The promise and success of lab-field generalizability in experimental economics: A reply to Levitt and List. In *The Methods of Modern Experimental Economics*, edited by Guillaume R. Frechette and Andrew Schotter. New York, NY: Oxford University Press, Chapter 14, 249-295.

Carter, John R., and Michael D. Irons. 1991. Are economists different, and if so, why? *The Journal of Economic Perspectives* 5 (2): 171-177.

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): 88-95.

Choo, C.Y. Lawrence, Miguel A. Fonseca, and Gareth D. Myles. 2016. Do students behave like real taxpayers in the lab? Evidence from a real effort tax compliance experiment. *Journal of Economic Behavior & Organization* 124: 102-114.

Cialdini, Robert B., Raymond R. Reno, and Carl A. Kallgren. 1990. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology* 58 (6): 1015-1026.

Coursey, Don L., Hovis, John L., and Schulze, William D. 1987. *The Quarterly Journal of Economics* 102 (3): 679-690.

Cowell, Frank. 1990. *Cheating the Government: The Economics of Evasion*. Cambridge, MA: The MIT Press.

Dawes, Robyn M. 1980. Social dilemmas. *Annual Review of Psychology* 31 (2): 169-193.

Dwenger, Nadja, Henrik Jacobsen Kleven, Imran Rasul, and Johannes Rincke. 2016. Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. *American Economic Journal: Economic Policy* 8 (3): 203-232.

Elster, Jon. 1989. *The Cement of Society – A Study of Social Order*. Cambridge, UK: Cambridge University Press.

Falk, Armin, and James J. Heckman. 2009. Lab experiments are a major source of knowledge in the social sciences. *Science* 326 (5952): 535-538.

Frechette, Guillaume R. 2015. Laboratory experiments: Professionals versus students. In *The Methods of Modern Experimental Economics*, edited by Guillaume R. Frechette and Andrew Schotter. New York, NY: Oxford University Press, Chapter 17, 360-390.

Frey, Bruno. 1997. *Not Just For the Money – An Economic Theory of Personal Motivation*. Cheltenham, United Kingdom: Edward Elgar Publishing Limited.

Gneezy, Uri, and John A. List. 2006. Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 74 (5): 1365-1384.

Hallsworth, Michael. 2014. The use of field experiments to increase tax compliance. *Oxford Review of Economic Policy* 30 (4): 658-679.

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

Harrison, Glenn W., Morten Lau, and E. Elisabet Rutström. 2015. Theory, experimental design and econometrics are complementary (and so are lab and field experiments). In *The Methods of Modern Experimental Economics*, edited by Guillaume R. Frechette and Andrew Schotter. New York, NY: Oxford University Press, Chapter 15, 296-338.

Harrison, Glenn W., and John A. List. 2004. Field experiments. *The Journal of Economic Literature* 42 (2): 1009-1055.

Internal Revenue Service Oversight Board. 2012. IRS Oversight Board Taxpayer Attitude Survey 2011. Washington, D.C.

Irwin, Kyle, and Brent Simpson. 2013. Do descriptive norms solve social dilemmas? Conformity and contributions in collective action groups. *Social Forces* 91: 1057-1084.

Iyer, Govind S., Philip M. J. Reckers, and Debra L. Sanders. 2010. Increasing tax compliance in Washington State: A field experiment. *National Tax Journal* 63 (1): 7-32.

Kagel, John H. 2015. Laboratory experiments: The lab in relationship to field experiments, field data, and economic theory. In *The Methods of Modern Experimental Economics*, edited by Guillaume R. Frechette and Andrew Schotter. New York, NY: Oxford University Press, Chapter 16, 339-359.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47 (2): 263-292.

Kastlunger, Barbara, Erich Kirchler, Luigi Mittone, and Julia Pitters. 2009. Sequences of audits, tax compliance, and taxpaying strategies. *Journal of Economic Psychology* 30 (3): 405-418.

Kessler, Judd B., and Lise Vesterlund. 2015. The external validity of laboratory experiments: The misleading emphasis on quantitative effects. In *The Methods of Modern Experimental Economics*, edited by Guillaume R. Frechette and Andrew Schotter. New York, NY: Oxford University Press, Chapter 18, 391-406.

Kirchler, Erich. 2007. *The Economic Psychology of Tax Behavior*. Cambridge, UK: Cambridge University Press.

Kirchler, Erich, Elizabeth Hoelzl, and Ingrid Wahl. 2008. Enforced versus voluntary tax compliance: The "slippery slope" framework. *Journal of Economic Psychology* 29 (2): 210-225.

Ledyard, John O. 1995. Public goods: A survey of experimental research. In *The Handbook of Experimental Economics*, edited by John H. Kagel and Alvin E. Roth. Princeton, NJ: Princeton University Press, 111-194.

Lerner Jennifer S., Roxana M. Gonzalez, Deborah A. Small, and Baruch Fischhoff. 2003. Effects of fear and anger on perceived risks of terrorism: A national field experiment. *Psychological Science* 14: 144-150.

Levitt, Steven D., and John A. List. 2007a. What do laboratory experiments measuring social

preferences reveal about the real world? *The Journal of Economic Perspectives* 21 (2): 153-174.

Levitt, Steven D., and John A. List. 2007b. Viewpoint: On the generalizability of lab behavior to the field. *Canadian Journal of Economics* 40 (2): 347–370.

Lewis, A. 1982. *The Psychology of Taxation*. Oxford, UK: Martin Robertson.

List, John A. 2003. Does Market Experience Eliminate Market Anomalies? *The Quarterly Journal of Economics* 118 (1): 41–71.

Maciejovsky, Boris, Erich Kirchler, and Herbert Schwarzenberger. 2007. Misperceptions of chance and loss repair: On the dynamics of tax compliance. *Journal of Economic Psychology* 28 (6): 678-691.

Maniadis, Zacharias, Fabio Tufano, and John A. List. 2014. One swallow doesn't make a summer: New evidence on anchoring effects. *The American Economic Review* 104 (1): 277-290.

McBarnet, Doreen. 2004. *Crime, Compliance, and Control*. Burlington, VT: Ashgate/Dartmouth Publishers Ltd.

Messer, Kent D., Harry M. Kaiser, and William D. Schulze. 2008. The problem of free riding in voluntary generic advertising: Parallelism and possible solutions from the lab. *American Journal of Agricultural Economics* 90 (2): 13.

Messer, Kent D., Homa Zarghamee, Harry M. Kaiser, and William D. Schulze. 2007. New hope for the voluntary contributions mechanism: The effects of framing and context. *Journal of Public Economics* 91 (9): 1783-1799.

Mittone, Luigi. 2006. Dynamic behaviour in tax evasion: An experimental approach. *The Journal of Socio-Economics* 35 (5): 813-835.

Park, Hee Sun, and Sandi W. Smith. 2007. Distinctiveness and influence of subjective norms, personal descriptive and injunctive norms, and societal descriptive and injunctive norms on behavioral intent: A case of two behaviors critical to organ donation. *Human Communication Research* 33: 194-218.

Rabin, Matthew. 2000. Risk aversion and expected-utility theory: A calibration theorem. *Econometrica* 68 (5): 1281-1292.

Reinganum, Jennifer F., and Louis L. Wilde. 1985. Income tax compliance in a principal-agent framework. *Journal of Public Economics* 26 (1): 1-18.

Reinganum, Jennifer F., and Louis L. Wilde. 1986. Equilibrium verification and reporting policies in a model of tax compliance. *International Economic Review* 27: 739-760.

Schmölders, Gunter. 1960. *Das Irrationale in der Offentlichen Finanzwirtschaft* [The Irrational in Public Finance]. Frankfurt am Main, Germany: Suhrkamp.

Schulze, William D., and Lee Hoffer. 2014. Norms in military environments. In *The Context of Military Environments – An Agenda for Basic Research on Social and Organizational Factors Relevant to Small Units*. Washington, D.C.: National Academies Press, Chapter 2, 23-40.

Schulze, William D., and Brian Wansink 2012. Toxics, toyotas, and terrorism: The behavioral economics of fear and stigma. *Risk Analysis* 32 (4): 678-694.

Schultz, Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2007. The constructive, destructive, and reconstructive power of social norms. *Psychological Science* 18 (5): 429-434.

Scotchmer, Suzanne, and Joel Slemrod. 1989. Randomness in tax enforcement. *Journal of Public Economics* 38 (1): 17-32.

Slemrod, Joel, and Caroline Weber. 2012. Evidence of the invisible: Toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance* 19 (1): 25-53.

Slemrod, Joel, and Shlomo Yitzhaki. 2002. Tax avoidance, evasion, and administration. In *Handbook of Public Economics, Volume 3*, edited by Alan J. Auerbach and Martin Feldstein. Amsterdam, London, and New York: Elsevier, 1423-1470.

Slovic, Paul, and Ellen Peters. 2006. Risk perception and affect. *Current Directions in Psychological Science* 15 (6): 322-325.

Tang, Thomas Li-Ping, and Yuh-Jia Chen. 2008. Intelligence versus wisdom: The love of money, Machiavellianism, and unethical behavior across college major and gender. *Journal of Business Ethics* 82 (1): 1-26.

Torgler, Benno. 2004. Moral-suasion: An alternative tax policy strategy? Evidence from a controlled field experiment in Switzerland. *Economics of Governance* 5 (3): 235-253.

Torgler, Benno. 2012. A field experiment on moral-suasion and tax compliance focusing on under-declaration and over-deduction. QUT School of Economics and Finance Working Paper No. 285. Queensland, Australia.

Vossler, Christian A., and Michael McKee. 2015. Behavioral effects of tax withholding on tax compliance: Implications for information initiatives. Department of Economics Working Paper 15-12, Appalachian State University. Boone, NC.

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

Wenzel, Michael. 2005. Misperception of social norms about tax compliance: From theory to intervention. *Journal of Economic Psychology* 26 (6): 862-883

Wenzel, Michael. 2006. A letter from the tax office: Compliance effects of informational and interpersonal justice. *Social Justice Research* 19 (3): 345-364.

Wenzel, Michael, and Natalie Taylor. 2004. An experimental evaluation of tax-reporting schedules: A case of evidence-based tax administration. *Journal of Public Economics* 88 (12): 2785-2799.

Figure 1. Withholding Amount in Control Treatment
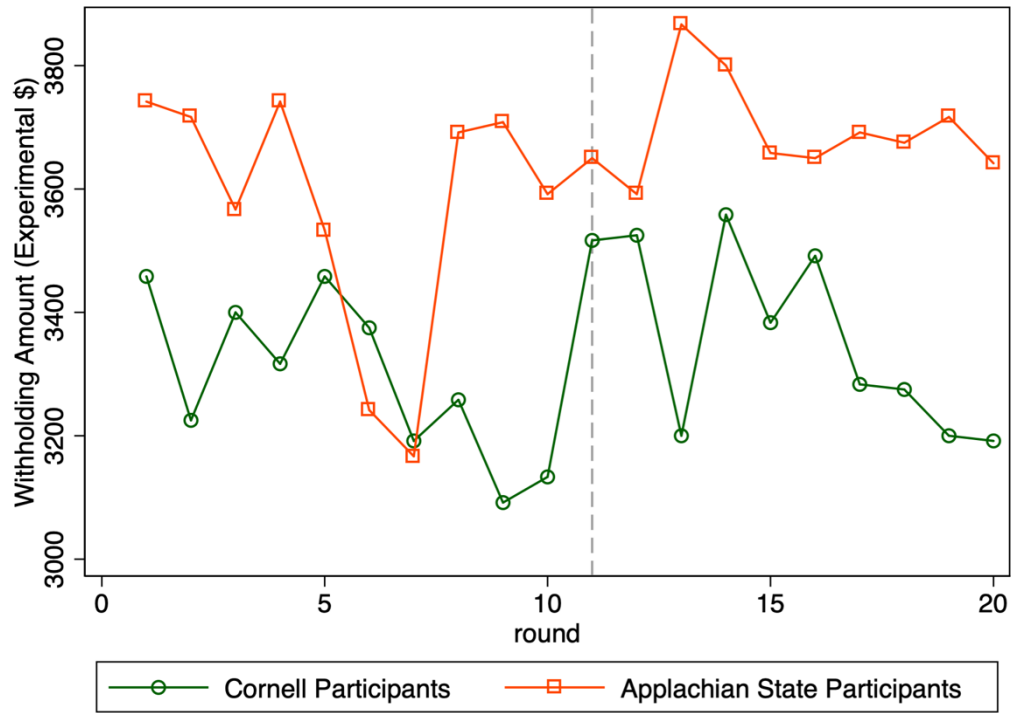
Figure 2. Tax Payment in Control Treatment
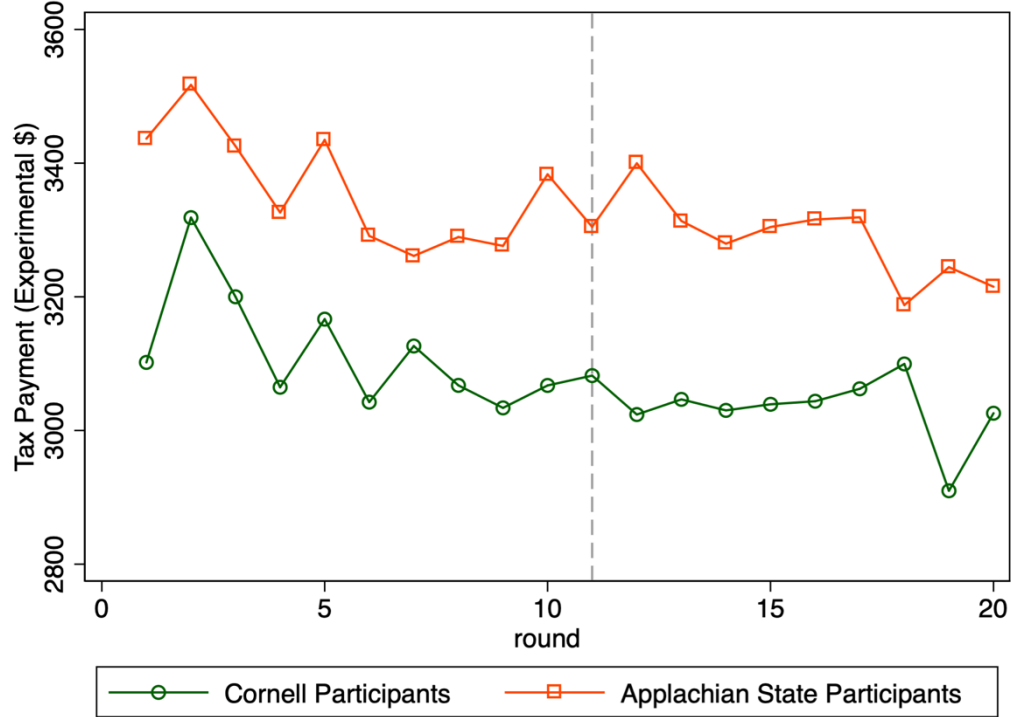
Table 1. Demographic Composition of Subjects by Site

	Cornell University	**Appalachian State University**
N	240	239
Gender		
Male	38.08%	43.04%
Female	61.92%	56.96%
Race		
White	54.62%	93.59%
Black or African American	10.92%	3.85%
American Indian or Alaska Native	--	0.43%
Asian	34.45%	2.14%
Income		
Less than $10,000	60.93%	87.44%
$10,000 - $20,000	11.63%	9.77%
$20,000 - $30,000	6.05%	2.69%
$30,000 - $40,000	5.58%	0.45%
$40,000 - $50,000	3.72%	--
$50,000 - $60,000	4.19%	--
$60,000 - $70,000	2.33%	--
$70,000 - $80,000	1.86%	--
$80,000 - $90,000	2.33%	--
$90,000 - $100,000	0.47%	--
$100,000 - $110,000	0.47%	--
$110,000 - $120,000	0.47%	--

Table 2. Simple Descriptive Statistics

Treatment	**Reported Taxes**			**Withholding Amount**		
	1st Half	**2st Half**	**Difference**	**1st Half**	**2nd Half**	**Difference**
T1: Control	3241.53 (20.64)	3162.33 (21.27)	79.19***	3430.42 (45.70)	3528.33 (46.31)	-97.92
T2: Descriptive Norm	3284.38 (22.43)	3149.00 (22.53)	135.37***	3302.52 (46.48)	3270.59 (47.01)	31.93
T3: Injunctive Norm, Approval-framed	3214.26 (21.82)	3214.26 (21.95)	68.53**	3534.58 (43.51)	3649.58 (44.00)	-115.00*
T4: Injunctive Norm, Disapproval-framed	3279.44 (21.58)	3279.44 (20.47)	63.39**	3383.75 (45.95)	3240.83 (45.96)	142.92**

Notes: Standard errors are in parentheses. t-tests on the difference between the first halves and second halves are conducted. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3. Regression Results: Determinants of *Reported Taxes*

Explanatory Variable	Specification			
	(1)	**(2)**	**(3)**	**(4)**
Tax Obligation	0.927***	0.923***	0.955***	0.941***
	(0.015)	(0.016)	(0.016)	(0.019)
Round	-13.504***	-11.957***	-12.169***	-13.470***
	(1.649)	(1.135)	(1.698)	(2.048)
Treatment 2 Second Half Dummy Variable (Descriptive Norm)	4.417	6.634	9.068	19.030
	(25.295)	(24.881)	(24.539)	(28.940)
Treatment 3 Second Half Dummy Variable (Injunctive Norm, Approval-framed)	65.055***	64.434**	63.373***	80.940**
	(29.489)	(29.174)	(29.134)	(35.800)
Treatment 4 Second Half Dummy Variable (Injunctive Norm, Disapproval-framed)	71.317***	67.198**	69.585***	87.130***
	(28.226)	(27.947)	(27.734)	(33.120)
Penalty		-0.023*	-0.241*	-0.0235*
		(0.012)	(0.013)	(0.013)
Public Good Benefit		0.016**	0.017**	0.0223**
		(0.008)	(0.007)	(0.001)
Withholding Enough (=1 if withholding amount is larger than tax liability; 0 otherwise)			101.793***	120.500***
			(23.591)	(29.300)
Withholding Amount			-0.003	-0.0104
			(0.010)	(0.012)
Constant	-275.379***	-347.362***	-504.547***	-643.687***
	(62.963)	(70.747)	(74.032)	(90.000)
Subject Fixed Effects?	Yes	Yes	Yes	Yes
R Squared	0.230	0.230	0.246	0.279
Observations	9572	9093	9093	6837

Notes: The dependent variable is *Reported Taxes*. Clustered standard errors at the subject level are in parentheses.
$^{*} p < 0.10$, $^{**} p < 0.05$, $^{***} p < 0.01$.

Table 4. Regression Results: Determinants of *Unreported Income*

Explanatory Variable	Specification		
	(1)	**(2)**	**(3)**
Tax Obligation	0.071***	0.075***	0.0448***
	(0.015)	(0.015)	(0.016)
Round	13.470***	11.920***	12.130***
	(1.633)	(1.698)	(1.679)
Treatment 2 Second Half Dummy Variable (Descriptive Norm)	-5.304	-7.236	-9.597
	(24.910)	(24.490)	(24.140)
Treatment 3 Second Half Dummy Variable (Injunctive Norm, Approval-framed)	-65.810**	-65.110**	-64.010**
	(29.280)	(28.940)	(28.890)
Treatment 4 Second Half Dummy Variable (Injunctive Norm, Disapproval-framed)	-70.940**	-67.090**	-69.510**
	(27.820)	(27.500)	(27.290)
Penalty		0.0242**	0.0249**
		(0.012)	(0.012)
Public Good Benefit		-0.016**	-0.0172**
		(0.008)	(0.008)
Withholding Enough (=1 if withholding amount is larger than tax liability; 0 otherwise)			-97.089***
			(23.059)
Withholding Amount			0.0019
			(0.010)
Constant	289.800***	364.800***	518.100***
	(61.640)	(69.350)	(72.870)
Subject Fixed Effects?	Yes	Yes	Yes
R Squared	0.014	0.017	0.047
Observations	9572	9093	9093

Notes: The dependent variable is *Unreported Income*. Clustered standard errors at the subject level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5. Regression Results: Heterogeneous Effects on *Reported Taxes*

	Gender	
Explanatory Variable	**Male**	**Female**
Treatment 2 Second Half Dummy Variable (Descriptive Norm)	13.020 (32.440)	0.420 (27.490)
Treatment 3 Second Half Dummy Variable (Injunctive Norm, Approval-framed)	47.740 (34.150)	76.150*** (25.940)
Treatment 4 Second Half Dummy Variable (Injunctive Norm, Disapproval-framed)	85.090** (35.730)	62.790** (25.260)
Observations	3919	5653
	Race	
Explanatory Variable	**White**	**Non-white**
Treatment 2 Second Half Dummy Variable (Descriptive Norm)	-16.470 (24.380)	59.630 (39.320)
Treatment 3 Second Half Dummy Variable (Injunctive Norm, Approval-framed)	64.840*** (24.320)	67.220* (39.410)
Treatment 4 Second Half Dummy Variable (Injunctive Norm, Disapproval-framed)	62.800*** (24.190)	93.780** (39.740)
Observations	6975	2597
	Religion	
Explanatory Variable	**Christian**	**Other**
Treatment 2 Second Half Dummy Variable (Descriptive Norm)	0.820 (27.980)	5.728 (31.020)
Treatment 3 Second Half Dummy Variable (Injunctive Norm, Approval-framed)	60.560* (30.980)	71.160** (27.960)
Treatment 4 Second Half Dummy Variable (Injunctive Norm, Disapproval-framed)	28.260 (29.930)	110.600*** (28.630)
Observations	4718	4854
	Volunteering Behavior	
Explanatory Variable	**Below Median**	**Above Median**
Treatment 2 Second Half Dummy Variable (Descriptive Norm)	-5.606 (29.740)	14.130 (28.900)
Treatment 3 Second Half Dummy Variable (Injunctive Norm, Approval-framed)	86.040*** (30.350)	51.230* (28.390)
Treatment 4 Second Half Dummy Variable (Injunctive Norm, Disapproval-framed)	25.470 (28.250)	122.300*** (30.340)
Observations	4639	4933

Notes: The dependent variable is *Reported Taxes*. The specification is the same as the main regression. Only the three treatment variables are reported. Clustered standard errors at the subject level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6. Regression Results: Determinants of *Withholding Amount*

Explanatory Variable	Specification	
	(1)	**(2)**
Tax Obligation	-0.019	-0.025
	(0.032)	(0.033)
Round	0.332	2.387
	(4.072)	(4.236)
Treatment 2 Second Half Dummy Variable (Descriptive Norm)	-35.348	-41.230
	(80.547)	(79.606)
Treatment 3 Second Half Dummy Variable (Injunctive Norm, Approval-framed)	118.182	96.206
	(78.727)	(78.811)
Treatment 4 Second Half Dummy Variable (Injunctive Norm, Disapproval-framed)	-146.226	-153.842*
	(91.636)	(91.567)
Penalty		-0.037
		(0.034)
Public Good Benefit		-0.028
		(0.022)
Constant	3497.750***	3608.431***
	(145.174)	(176.073)
Subject Fixed Effects?	Yes	Yes
R Squared	0.005	0.004
Observations	9572	9093

Notes: The dependent variable is *Withholding Amount*. Clustered standard errors at the subject level are in parentheses. $^{*} p < 0.10$, $^{**} p < 0.05$, $^{***} p < 0.01$.

TAX EXPERIMENT INSTRUCTIONS

Welcome to the Laboratory for Experimental Economics and Decision Research (Appalachian Experimental Economics Laboratory). This experiment is about tax compliance. Note that deception is NOT allowed in economics experiments. You will be compensated in cash for your participation at the end of the experiment. The amount you receive is based on choices you make during the experiment. If you have any questions during the experiment, please raise your hand and someone will come to assist you.

Your income in each round will be determined in two ways. First, your certain income in all of the rounds will be determined by estimating the number of gumballs in the one-quart jar at the front of the lab. If you estimate the number correctly, you will receive 10,000 experimental dollars of certain income in each round. If you do not exactly estimate the number of gumballs, your certain income will be reduced by 50 times your error in the number of gumballs in the jar. So, for example, if your estimate is off by 10 gumballs, your certain income in EACH round would be 10,000 – 10*50 = 9,500 experimental dollars. You are guaranteed a minimum certain income of 5,000 experimental dollars even if your error is more than 100 gumballs. The second component of your income is determined by a random draw on each round. In addition to the certain income that you will receive in each round, you will receive a random income of between 5,000 and 10,000 experimental dollars. Any dollar amount in this range is equally likely. Each person in the experiment will get a different random draw from the computer. This random component is meant to simulate the uncertainty most people face in estimating their annual total income due to uncertainty over the size of a possible salary or wage increase, annual bonus, possible gifts or inheritance, overtime, etc. Note that the highest income that you could make in a round is 10,000 experimental dollars in certain income, if you exactly estimated the number of gumballs, plus an additional 10,000 in random income.

The experiment involves three stages in each round. As in the actual tax system, in Stage 1, you will be asked to choose a withholding amount WITHOUT knowing your exact actual income. After you choose your withholding amount, you will learn your actual income, and in Stage 2 you will file your tax return where you can claim deductions that will reduce your tax obligation.

After you file your tax return, any tax you owe in addition to the amount withheld will be deducted from your balance with a 20% penalty from the tax authority on unpaid taxes (under-withholding), or, if you withheld too much, your refund will be calculated and returned to you. Your income that you choose NOT to withhold will earn interest on that amount of 10% since, in an actual tax decision, you could put the money you keep during the tax year in the bank and earn interest or enjoy the use of it during the year. Also, based on taxes collected, each participant will receive a tax-based benefit equal to 1.4 times the total of taxes collected (including the collected unpaid tax but not including the penalty. See more below), divided by the number of participants. That is, we will multiply the total collected tax by 1.4 and distribute the amount equally among all participants. This benefit will be paid based on collections in that round. In addition each participant will receive their after tax (and penalty, see below) income at the end of each session.

In Stage 3, there is a possibility that your tax return will be audited. Your odds of being audited are 1/25 (or 4%). The chance does not increase or decrease depending on your current or past reporting choices or on the decision made by others in the group. In addition, the random draw is independent in every round. This is a random selection process where, in Stage 3, the computer will randomly determine which participant(s), if any, will be confidentially audited on that round.

If you are selected for an audit, your tax payments will be checked against the taxes owed. If you underpaid your taxes by claiming too many deductions, all unpaid taxes will be discovered. If you are not audited; however, no unpaid taxes will be discovered. If audited, you will have to pay back any taxes owed for that round because you claimed too much in deductions. Unpaid taxes are calculated as the difference between your actual and reported amounts multiplied by the tax rate. Any unpaid taxes discovered in the audit must be paid back. If you have unpaid taxes, a penalty of 100% will also be assessed. What this means is that, if you are audited, for every lab dollar in unpaid taxes, you will have to pay back the 1 lab dollar you owe in taxes and in addition 1 lab dollar in penalty.

At the end of the experiment all of your earnings in the experiment will be added for all of the rounds. You will be given cash equal to $1 for every 15,000 experimental dollars you earn. The number of rounds is predetermined but you will not be told how many rounds are in the experiment.

Please remember that taxes collected do provide benefits, so in this experiment, all of the money collected through taxes, in each session will be multiplied by 1.4 by the experimenters and returned to participants in the experiment in equal shares. This tax-based benefit payment will be added to the payoff at the end of each round.

We will begin with some practice rounds so you can see how the experiment works.

Please raise your hand if you have a question.

Please look at your computer screen and follow the instructions and answer the questions on the screen.

Figure A1. Guessing Task Screen

Remember, your certain income in *ALL* of the rounds will be determined by estimating the number of gumballs in the one-quart jar at the front of the lab. If you estimate the number correctly, you will receive 10,000 experimental dollars of certain income in each round. If you do not exactly estimate the number of gumballs, your certain income will be reduced by fifty times your error in the number of marbles in the jar.

Please enter your estimate in the yellow box then click the PROCEED button.

Your estimate of gumballs in the jar is:

PROCEED

Figure A2. Withholding Decision Screen

Determine Witholding

Round   1

Your certain income in this round is: $7,000

The range of your random income is: $5000-$10000

Please select your allowance to determine your withholding in this round (Tax Rate = 25%)

	Allowances	Tax Withheld
○	0	$6,000
○	1	$5,000
○	2	$4,000
○	3	$3,000
○	4	$2,000
○	5	$1,000
○	6	$500

There is a 20% penalty for under-withholding taxes plus you will have to pay any taxes owed. However, if you decided to keep the money you under-withheld for the tax year in your bank account it would earn interest or provide enjoyment. To simulate that, we will add 10% of any under-withholding to your account on each round.

SUBMIT

Figure A3. Tax Filing Decision Screen

Figure A4. Outcome Screen

Please answer the following questions to the best of your knowledge. All information is completely anonymous and confidential.

The first questions relate to your experience in today's experiment.

1) How well were you compensated for your time and effort in this experiment?

(poorly compensated)   1   2   3   4   5 (well compensated)

2) How well do you feel you understood the instructions for the experiment?

(poorly understood)   1   2   3   4   5 (well understood)

3) Have you previously participated in a paid economics experiment before?
   1. Yes
   2. No

The next questions tell us something about you.

4) What is your age? [number of years old (18 – 90+)]

5) What is your gender?
   1. Male
   2. Female

6) Are you Spanish/Hispanic/Latino?
   1. Yes
   2. No

7) What is your race?
   1. White
   2. Black or African-American
   3. American Indian or Alaska Native
   4. Asian
   5. Native Hawaiian or other Pacific Islander

8) What is your religion?
   1. Christian (Which denomination? _____)
   2. Jewish
   3. Muslim
   4. Buddhist
   5. Hindu
   6. Other (Please specify _____)
   7. I don't follow any specific religion, but I do have spiritual beliefs (for example, you believe there is some other power or force outside yourself which might influence your life).
   8. None

9) About how often, if ever, have you attended religious services in the last year?
   1. Once a week or more
   2. Two or three times a month

      3. Once a month
      4. A few times a year or less
      5. Never
      6. Not Applicable

6) Please indicate your marital status:
      1. Single
      2. Married
      3. Widowed
      4. Divorced

7) Are you currently a student?
      1. Yes (go to question 7a)
      2. No (skip to question 8)

7a) Which best describes your student status?
      1. Full-time
      2. Part-time

7b) What year are you at school?
      1. Freshman
      2. Sophomore
      3. Junior
      4. Senior
      5. Graduate student

7c) What is your intended or declared major? _____

8) What is your <u>highest</u> level of education?
      1. Completed Graduate Degree
      2. Some Post-graduate Study
      3. Completed Undergraduate Degree
      4. Some College or University
      5. Trade School or Vocational Training
      6. Completed High School
      7. Did not graduate from high school

9) Which of the following best describes your current employment status?
      1. Employed full-time
      2. Employed part-time
      3. Self-employed full-time
      4. Self-employed part-time
      5. Unemployed (skip to question 11)
      6. Retired (skip to question 11)

10) In what industry do you work? [Pull down menu or radio buttons.]
      1. Construction
      2. Manufacturing
      3. Retail Trade
      4. Finance, Insurance, Real Estate
      5. Education (including private, state, and local government schools)

6. Health Care and Social Assistance (including private, state, and local government hospitals)
7. Arts, Entertainment, and Recreation
8. Accommodation (including hotels and motels)
9. Food Services (including restaurants, cafes and bars)
10. Federal, State, and Local Government (excluding state and local schools and hospitals)
11. Other

11) Please indicate your total <u>personal</u> income for last year. [Use pull down with ranges in $10K intervals. Max level > 200K]

12) Please indicate the degree to which the following statements describe you:
(for each statement, have a scale of 1 – 5, where 1 is "Does NOT describe me well" and 5 is "Describes me very well")
1. I often have tender, concerned feelings for people less fortunate than me.
2. Sometimes I don't feel very sorry for other people when they are having problems.
3. When I see someone being taken advantage of, I feel kind of protective towards them.
4. Other people's misfortunes do not usually disturb me a great deal.
5. When I see someone being treated unfairly, I sometimes don't feel very much pity for them.
6. I am often quite touched by things that I see happen.
7. I would describe myself as a pretty soft-hearted person.

13) During the past 12 months, how often have you done each of the following:
1. Donated blood
   a. More than 3 times in the past year
   b. At least 2 or 3 times in the past year
   c. Once in the past year
   d. Not at all in the past year
   e. Don't know
   f. Prefer not to respond
2. Done volunteer work
   a. More than 3 times in the past year
   b. At least 2 or 3 times in the past year
   c. Once in the past year
   d. Not at all in the past year
   e. Don't know
   f. Prefer not to respond
3. Given food or money to a homeless person
   a. More than 3 times in the past year
   b. At least 2 or 3 times in the past year
   c. Once in the past year
   d. Not at all in the past year
   e. Don't know
   f. Prefer not to respond
4. Given money to charity
   a. More than 3 times in the past year
   b. At least 2 or 3 times in the past year
   c. Once in the past year
   d. Not at all in the past year
   e. Don't know
   f. Prefer not to respond

---

[1] See Hallsworth (2014) for a comprehensive survey of this literature.

[2] See especially Harrison and List (2004) and Levitt and List (2007a, b) for discussions of the advantages of field experiments in economic research.

[3] See Falk and Heckman (2009), Camerer (2015), Frechette (2015), Harrison, Lau, and Rutström (2015), Kagel (2015), and Kessler and Vesterlund (2015) for detailed discussions of the limitations of field experiments, along with both the advantages of laboratory experiments and the complementarities of laboratory and field experiments.

[4] See Levitt and List (2007a, b) for a general critique of laboratory experiments. For robust responses to this critique, see Falk and Heckman (2009), Camerer (2015), Frechette (2015), Harrison, Lau, and Rutström (2015), Kagel (2015), and Kessler and Vesterlund (2015). Also, see Alm, Bloomquist, and McKee (2015) for specific evidence on the external validity of tax compliance experiments, who find that student and non-student behaviors are similar.

[5] For an alternative and contrary view on the external validity of laboratory experiments, see Choo, Fonseca, and Myles (2016).

[6] We are grateful to an anonymous referee for suggesting this alternative approach.

[7] We do not test for risk aversion for four reasons. First, because of the relatively small stakes used in our experiment, risk aversion is inconsistent with expected utility theory, as demonstrated by Rabin (2000). He theoretically analyzes risk aversion for the small stakes used in most laboratory experiments, and concludes that "…expected-utility theory is manifestly not close to the right explanation of risk attitudes over modest stakes", such as those that apply in most all experimental studies. Indeed, Rabin (2000) concludes that "…within the expected-utility model, anything but virtual risk neutrality over modest stakes implies manifestly unrealistic risk aversion over large stakes". Second, from the perspective of behavioral economics, testing for risk aversion is problematic because, relative to the reference point, many individuals are risk-averse or risk-neutral in gains but risk-seeking in losses (Kahneman and Tversky 1979). As a result, risk preference may depend on whether decisions are framed as gains or losses, and methods for testing risk aversion do not typically test for framing effects. Third, our experiment involves many rounds, and risk preferences may change over rounds since repeated market experience can induce framing effects (Coursey, Hovis, and Schulze 1987; List 2003). Again, however, methods for testing risk aversion do not typically consider that risk preferences might change with experience (or with framing effects). Finally, the weighting function of prospect theory predicts overweighting of low probabilities and underweighting of high probabilities, and such behavior would give the appearance of risk aversion for low probabilities and risk seeking for high probabilities. Once again, methods for testing risk aversion do not typically consider that risk preferences might change with the weighting of high and low probabilities. For all of these reasons, a single measure of risk aversion is not likely to be a useful concept for our laboratory experiment since our experiment has modest stakes, includes both gains and losses, employs multiple rounds with changing subject experience, and uses probabilities of audit that imply different weights relative to actual probabilities of audit.

[8] See Cowell (1990), Andreoni, Erard, and Feinstein (1998), Slemrod and Yitzhaki (2002), and Alm (2012, 2019) for comprehensive surveys of this literature.

[9] See Vossler and McKee (2015, 2017) for more detailed discussions of these considerations, from which much of our framework is derived. Note also that this simple framework assumes that the various fiscal parameters do not depend on the taxpayer's decisions. In particular, the

audit probability is assumed to be fixed and independent of any information that the taxpayer conveys to the tax authority on the tax return. However, individuals do not typically face a fixed and independent, or "exogenous", probability of audit. Instead, the tax authority often uses information from the tax returns to determine strategically whom to audit, so that the audit probability is "endogenous", dependent in part on the behavior of the taxpayer and the tax authority. See Reinganum and Wilde (1985, 1986) for analyses of such endogenous audit selection rules. This assumption is made largely for convenience, and it does not affect our main results.

[10] See Alm (1988), Scotchmer and Slemrod (1989), and Alm, Jackson, and McKee (1992) for analyses of the effects of different aspects (and their potentially conflicting effects) of uncertainty on taxpayer compliance.

[11] For example, Schmölders (1960) argues that individuals have an *intrinsic motivation* to cooperate. Frey (1997) suggests that compliance depends upon an individual's *tax morale*. Kirchler, Hoelzl, and Wahl (2008) explore the interaction between enforcement effort (*power*) and facilitation (*trust*) on the part of the tax authority via a *slippery slope* framework. McBarnet (2004) suggests that people may choose to comply willingly (what she terms *committed compliance*), they may choose to comply unwillingly (*capitulative compliance*), they may take full advantage of the law in minimizing their taxes (*creative compliance*), or they may choose noncompliance. Braithwaite (2009) argues that individuals differ in their motivations to comply with tax law, and she identifies both positive (*deference*) and negative (or *defiance*) motivations. There are of course still other approaches.

[12] Note that this consideration helps to motivate the importance of withholding. If taxes have already been withheld, then there is no pain of payment at the time of tax filing; however, if a taxpayer has not withheld enough, then they suffer a loss, which makes them more motivated to evade taxes.

[13] See Alm, Clark, and Leibel (2016) and, especially, Vossler and McKee (2015, 2017) for detailed analyses of similarly structured models.

[14] More precisely, the individual is more likely to report all income, given the all-or-none nature of individual behavior.

[15] Note that these same basic hypotheses tend to emerge from a model in which the individual is assumed to maximize expected utility, not expected value. Admittedly, however, the presence of risk aversion makes the hypotheses from expected utility maximization ambiguous, since the comparative statics of the individual's choices now depend upon the individual's attitude toward risk.

[16] Note that reported *taxes* is a linear transformation of reported *income*, given that the tax rate on reported income is a constant.

[17] Alm, McClelland, and Schulze (1991) compared subject compliance behavior in experiments that used tax language and in experiments that used neutral terminology, and they found no difference in subject compliance behavior.

[18] The nature by which taxpayers understand the audit probability is critical to the theories of tax compliance based on the "economics of crime" approach to evasion. Alm, Jackson, and McKee (2009) investigate the information transmission process in detail.

[19] The multiplier of 1.4 is meant to capture the consumer surplus that public goods typically generate. For example, see Alm, McClelland, and Schulze (1992), who vary the size of the multiplier in different treatments.

[20] For example, see Carter and Irons (1991) and Tang and Chen (2008).

[21] As noted earlier, one student was a no-show for Treatment 2 at Appalachian State University.

[22] Also see Alm, Jackson, and McKee (1993), Alm, McClelland, and Schulze (1999), and Hallsworth et al. (2017).

[23] These social norm messages refer to the behavior of students in previous sessions of the experiment and to the behavior of U.S. taxpayers generally. The statements that refer to student behavior were based on three pilot sessions conducted at each university, during which participants were shown descriptive social norm messages referencing only nationwide behavior rather than university-specific behavior. The statistics that refer to the behavior of American taxpayers are drawn from various sources, as noted in the table, although the sources were not actually included in the table shown to participants due to space and formatting constraints.

[24] The statistics on withholding are based on the behavior of students during the pilot rounds of the experiment and on data on individual tax returns filed in 2014. See http://www.irs.gov/uac/Dec-26-2014.

[25] See IRS Oversight Board (2012).

[26] See Otto and Schultz (1990).

[27] See IRS Oversight Board (2012).

[28] See Otto and Schultz (1990).

[29] For example, see Mittone (2006), Maciejovsky, Kirchler, and Schwarzenberger (2007), and Kastlunger et al. (2009) for experimental studies that find a bomb-crater effect.

[30] Note that our use of a fixed effects model at the subject level (i.e., the treatment effect captured in the regression is from within-subject variation) includes any possible difference between Cornell University and Appalachian State University subjects.

[31] Note that we have also estimated a panel logit model for whether or not the individual evades any taxes, where the dependent variable is a dummy variable equal to 1 if the individual reported taxes less than the amount legally due and 0 otherwise. These estimation results are entirely consistent with the results for *Reported Taxes*; that is, all three treatment variables decrease the probability that an individual evades any taxes, and both appeals to injunctive norms have a large, negative, and statistically significant impact on the probability of evasion, with the magnitudes of the injunctive norm impacts virtually identical in size. These results are not reported but are available upon request.

[32] We are grateful to an anonymous referee for suggesting these additional robustness tests.

[33] Note that some other studies of social norm messaging have shown a significant decay of compliance over time or between messages, although these studies have not generally been in the context of tax compliance. For example, see Alcott (2011) for a field study of appeals to social norms in the context of energy conservation.