Journal of Behavioral and Experimental Economics 71 (2017) 45–55

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

The role of gender in the provision of public goods through tax compliance

David M. Bruner[a,*], John D'Attoma[b], Sven Steinmo[c]

[a] Department of Economics, Appalachian State University, Boone, NC 28608, USA
[b] Tax Administration Centre, University of Exeter Business School, Exeter EX4 4PU, UK
[c] Department of Political Science, University of Colorado, Boulder, CO 80309, USA

ARTICLE INFO

Article history:
Received 28 November 2016
Revised 6 August 2017
Accepted 8 September 2017
Available online 15 September 2017

JEL classification:
H2
H26
C91

Keywords:
Individual income tax
Public goods
Gender
Experiments

ABSTRACT

The existing experimental literature suggests women are more compliant than men when paying taxes but may free ride more when contributing to public goods. It is unclear which effect dominates when paying for public goods through taxation. Experiments conducted in three European countries and the U.S. are used to investigate this issue. The results suggest that women bear a greater burden of the provision of public goods for the parameters in the experiment. The results indicate the gender gap in compliance is due to differences in both the extensive and intensive margins.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Although economic theory is largely silent on the role that gender plays in decision-making, there is now considerable experimental evidence of gender differences in behavior across a broad range of economic environments. This paper explores whether there are gender differences in the propensity to contribute to the provision of public goods. To answer this question, however, it is important to recognize that in modern societies such public goods are financed through taxation. Hence, a closely related, albeit distinct question arises, are there gender differences in the willingness to pay taxes? The provision of public goods via tax compliance is a particularly complex choice to examine that involves attitudes towards altruism, fairness, honesty, obedience, risk, and trust; all of which are dimensions of preferences that seem to be shaped, at least in part, by gender (Croson and Gneezy, 2009). Hence it is reasonable, given the findings of previous studies, to expect willingness to pay taxes and the resulting provision of public goods to vary by gender. The direction in which behavior will differ is less obvious. On one hand, results from a variety of experiments, covering a diverse array of motivations, are consistent with women being more willing to pay taxes. On the other hand, experiments exploiting public goods games suggest men may be more willing to voluntarily contribute to public goods. Hence, existing evidence implies there are potentially offsetting gender effects when it comes to provision of public goods through taxation.

Experimental evidence of gender differences in risk taking, competitiveness, honesty and obedience, all of which are indicators of tax compliance, give credence to the notion that there may be gender differences in the willingness to pay taxes. First, studies of risk preference suggest women may be more risk averse than men (Eckel and Grossman, 2008; Eckel and Füllbrunn, 2015; Charness and Gneezy, 2012), although the majority of studies find no significant difference across genders (Filippin and Crosetto, 2016). Given the inherent uncertainty of the enforcement regime, a greater aversion to risk implies greater tax compliance. Second, there is evidence that men are more competitive than women because they are more overconfident (Niederle and Vesterlund, 2007, 2008, 2011). Such overconfidence could well translate into lower compliance when paying taxes if individuals underestimate their risk of audit. Third, there is some evidence that girls are more honest than boys (Bucciol and Piovesan, 2011) and women are more honest than men (Dreber and Johannesson, 2008); however, many experiments on lie aversion report

* Corresponding author.
E-mail address: brunerdm@appstate.edu (D.M. Bruner).

http://dx.doi.org/10.1016/j.socec.2017.09.001
2214-8043/© 2017 Elsevier Inc. All rights reserved.

no significant difference across genders (Cappelen et al., 2013; Childs, 2012; Gylfason et al., 2013). Still, to the extent that differences exist, an inclination towards honesty favors greater tax compliance. Finally, both Cadsby et al. (2006) and Karakostas and Zizzo (2016) provide evidence that women are more obedient to authority than men, even though the difference is only significant in the former study. Nonetheless, such a difference would suggest greater tax compliance for women than men. Accordingly, it should come as no surprise that several experiments report a gender compliance effect, where tax compliance is higher for female subjects (Alm et al., 2009, 2010, 2012; Spicer and Becker, 1980).

There are two important caveats regarding the present research question that should be noted. First, all of the tax compliance experiments that report a gender difference were conducted in the U.S. Therefore, it is possible that the gender difference is culturally biased. For instance, Alm and Torgler (2006) report significant differences in tax compliance behavior in experiments conducted in the U.S. and Spain. Public goods games have also demonstrated varying results depending on the country in which the experiment was conducted and the context of the game (Brown-Kruse and Hummels 1993; Solow and Kirkwood, 2002; Cadsby and Maynes 1998). Hence, behavioral differences in men and women can be due to differences in context and protocol (Croson and Gneezy, 2012, 463). More importantly, none of the tax compliance experiments that report a gender difference use the collected taxes to finance a public good. Thus the question remains, is there a gender difference in the willingness to contribute to public goods?

Several studies have attempted to address this question using experiments on public goods games. However, the results of these studies have been inconsistent. Both Brown-Kruse and Hummels (1993) and Solow and Kirkwood (2002) found evidence of a gender free-riding effect, where women contributed significantly less than men across all of their experimental treatments. On the other hand, Nowell and Tinkler (1994) reported results from a similar experiment that suggests women contribute more than men, if the group is composed entirely of females, which is an unlikely situation in the real-world. In between, Isaac et al. (1985), Stockard et al. (1988), and Cadsby and Maynes (1998) find no significant difference in contributions across genders.[1] So it seems if there is a gender difference, men are more willing to contribute to public goods.

The present study is the first attempt, to our knowledge, to bridge the gap between these two separate but related literatures by investigating whether gender differences exist in the willingness to contribute to public goods through tax compliance. To explore this issue, the results from laboratory experiments conducted across three countries in Europe (Italy, Sweden, and the U.K.) and the U.S. are utilized.[2] Subjects earned income in a real effort task and reported their income for tax purposes, which were then used to finance a public good. The size and scope of this study makes it the most comprehensive investigation to date. Furthermore, conducting experiments in multiple countries allows us to test for cultural differences in gender effects.

The key to our identification strategy to distinguish willingness to pay taxes from willingness to contribute to public goods is to observe tax compliance in both the absence and presence of a public good. As in the previously cited studies, pure tax compliance is observed when there is no public good. Observed behavior when there is a public good is a combination of tax compliance and willingness to contribute. Therefore, the change in an individual's tax compliance behavior associated with the introduction of a pubic good (i.e., increasing the multiplier from zero to a positive amount) captures their willingness to contribute to the public good.

Our results suggest overall women are more willing to pay taxes and men are more willing to contribute to public goods. The observed gender difference in compliance is robust across countries and consistent with the previous literature. Contrary to conventional wisdom, the gender gap in compliance does not appear to be due to differences in risk aversion. Moreover, the analysis adds to the existing literature by demonstrating this gender difference in compliance is due to differences in both the extensive and intensive margins; women are less likely to underreport their income and do so by smaller amounts. This finding is also consistent across all countries in the sample. While the willingness to contribute to public goods favors men in all countries, the difference is only significant in Italy and Sweden. This is consistent with previous evidence that men are more sensitive to the cost of generosity (Andreoni and Vesterlund, 2001; Cox and Deck, 2006). Still, for the parameter values used in the experiment, women contribute more to public goods that are financed through taxation.

The remainder of the paper is organized as follows. Section 2 presents the experimental design and formally constructs our behavioral hypotheses. Section 3 reports the results of the statistical analysis of behavior in the experiment. Finally, Section 4 discusses the results and their implications for policy.

2. Experimental design

2.1. Experimental protocols

Our experimental design implements the fundamental elements of any voluntary tax reporting system. Participants earn income by performing a clerical task, and self-report their earned income to a tax authority. Reported income is multiplied by the known tax rate to determine tax liability. To investigate the propensity to contribute to public goods, reported taxes are collected and placed into a group fund, which is multiplied by a known multiplier, and then divided equally among the participants in a session. At the end of an experimental session, reported income is subject to a random audit process, which performs without error. If an audit occurs, underreported income is discovered, and underpaid taxes as well as the associated penalties are collected. Net income is equal to earned income plus the share of the group fund less paid taxes and penalties, if applicable.

Experimental sessions consisted of 3 stages, with 3 income reporting rounds within each for a total of 9 income reporting rounds. Each of the stages implements changes in a specific experimental parameter. In the first stage, the public good multiplier, m, is varied, which is our primary focus in this analysis. The second stage varied the tax rate and the third stage varied the progressivity of the tax system. At the beginning of each stage, participants performed a clerical task which consisted of typing random characters into the computer. Participants were paid a piece rate of 10 tokens for each line they typed correctly; the sum constituted their income for the three reporting rounds in a stage. These tokens were exchanged for the domestic currency at a rate of 0.01 per token at the end of the experiment.

At the beginning of each income reporting round, participants were informed of the relevant experimental parameters. Specifically, they were informed that the tax rate was 30 percent (in the first three rounds) of reported income, the probability of an audit was 5 percent and the fine rate on unpaid taxes was 100 percent. In addition, they were informed of the multiplier, m, on the public good. That is, taxes were collected and summed, multiplied by $m \in \{0, 1, 2\}$, and then divided equally among the participants

[1] Both Andreoni and Vesterlund (2001) and Cox and Deck (2006) provide evidence that suggests these differences could be due to the cost of generosity.

[2] The basic experimental design has been utilized for more than 20 years (Alm et al., 1992). See Alm and McKee (1998) and Torgler (2002) for a review of the literature.

Table 1
Experimental design.

Stage 1			
Order	1	2	3
A	$m = 0$	$m = 1$	$m = 2$
B	$m = 2$	$m = 1$	$m = 0$

in a session.[3] In every case, the expected payoff is maximized by reporting zero income.[4] In most sessions, the multiplier was increased in the first three rounds; it was 0 in the first round, 1 in the second round, and 2 in the third round. As a robustness check additional sessions were conducted in Italy where the multiplier was decreased; it was 2 in the first round, 1 in the second round, and 0 in the third round. The experimental design is summarized in Table 1.

After the tax reporting rounds, subjects were tasked with 15 allocation decisions to elicit their 'social value orientation' (Murphy et. al., 2011). These decisions required participants to choose an allocation of tokens between themselves and an anonymous partner. These tokens were exchanged for the domestic currency at a rate of 0.0003 per token at the end of the experiment. The first six decisions, depicted in Table 2, were constructed to assess whether an individual had primarily individualistic or prosocial/altruistic motives.[5]

Finally, in the six sessions that varied the treatment orders, participant's risk attitudes were elicited using a multiple price list (Andersen et al., 2006).[6] The mechanism presents participants with a menu of 10 choices between a binary lottery and a constant sum of money, of 70 tokens, as shown in Table 3. Each binary lottery has a fixed high payout of 100 tokens and a fixed low payout of 40 tokens. These token amounts were chosen to approximate the binary choice between full compliance and full evasion of the median income earner in under order A sessions. Accordingly, these tokens were exchanged for domestic currency at a rate of 0.01 per token, as in the first round of the experiment. As participants proceeded through the menu the probability of the high (low) payout is increased (decreased) from 0.15 to 0.95 in increments of 0.1 to induce subjects to switch from the safe to the risky option.[7] The point at which a participant switches indicates their risk preference.

Upon the completion of the risk preference elicitation task, if administered, participants then proceeded to complete a survey. The survey collected demographic information about the participant as well as inquired about their potential motivations regarding their behavior in the experiment. After completion of the survey, participants were paid individually in private.

The experiment was conducted in 4 different countries to control for possible cultural bias: U.S., U.K., Sweden, and Italy. Furthermore, the experiment was conducted in at least two separate universities within each country.[8] Subjects were recruited by email via each lab's Online Recruitment System for Experimental Economics (ORSEE) (Greiner, 2015). The sessions were programmed and conducted with the software Z-Tree (Fischbacher, 2007), as well as with the software Behavery. Experimental sessions lasted approximately 90 minutes. Average earnings in the experiment were approximately $14 (USD), in addition to a $5 (USD) show-up fee.

2.2. Behavioral hypotheses

Since our primary interest is in the effect of gender on the provision of public goods through tax compliance, we restrict our hypotheses and analysis to the first three rounds of the experiment. Let the fraction of income reported by participant i in decision round j, Y_{ij}, be given by

$$Y_{ij} = \beta_0 + \beta_1 P + \beta_2 P^2 + \beta_3 F_j + \beta_4 PF_j + \beta_5 P^2 F_j + \boldsymbol{\beta X_j} + \varepsilon_{ij}, \quad (1)$$

where P is an indicator variable equal to one if the multiplier on the public good is equal to one; P^2 is an indicator variable equal to one if the multiplier on the public good is equal to two; F_j is an indicator variable equal to one if participant j is female; PF_j is an interaction variable equal to one if the multiplier on the public good is equal to one and participant j is female; P^2F_j is an interaction variable equal to one if the multiplier on the public good is equal to two and participant j is female; $\boldsymbol{X_j}$ is a set of demographic control variables for participant j; and ε_{ij} is an unobserved error term.

There are two fundamental hypotheses regarding the effects of gender based on previous findings from experiments on public goods and tax compliance. First, we test the hypothesis that compliance increases with the multiplier on the public good. We state our first hypothesis formally as follows:

Hypothesis 1: $\beta_2 > \beta_1 > 0$: *Increasing the multiplier from zero to two on the public good should increase compliance for both males and females.*

Second, there is growing and consistent evidence that suggests females have higher tax compliance. We state the second hypothesis formally as follows:

Hypothesis 2: $\beta_3 > 0$: *Female participants should have a higher tax compliance rate than male participants.*

Finally, although there are mixed results in the literature, there is evidence that males contribute more than females to public goods. This implies that increasing the public good multiplier should have a greater effect for males than females. We state our third hypothesis formally as follows:

Hypothesis 3: $\beta_5 < 0$ & $\beta_4 < 0$: *Increasing the public good multiplier should increase tax compliance more for males than females.*

[3] The marginal per capita return on the public good was $\frac{m}{N}$, where N denotes the number of subjects in a session. Due to variation in the show-up rates across sessions and sizes of the labs across universities N varied from 6 to 32.

[4] Given the tax rate of 30%, a 5% probability of audit, and a fine rate of 100% on unpaid taxes, the expected payoff of subject i is $E[\pi_i] = I_i - 0.3x_i + \frac{m}{N}0.3\sum_{i=1}^{N} x_i - 0.03(I_i - x_i)$, where I_i denotes earned income and x_i denotes reported income, depends on the values of m and N. For $N = 6$ (the minimum for our sample), the expected marginal benefit of reporting income is 0.03, 0.08, and 0.13 when m is 0, 1, and 2, respectively. These values decline as N increases. Hence, given the marginal cost is 0.3, the expected payoff is maximized when zero income is reported.

[5] The last nine decisions determined whether prosocial behavior was driven by inequality aversion or joint gain maximization. Since the motivation for prosocial behavior is not relevant for this research question, these decisions are omitted from Table 2 and the subsequent analysis.

[6] This mechanism was made popular by Holt and Laury (2002).

[7] The last decision involved a choice of 100% chance of winning versus the constant sum of money to ensure participants were paying attention.

[8] The experimental sites included Bologna Laboratory for Experiments in Social Sciences, Centro d'Economia Sperimentale A Roma Est, and Experimental Economics Lab of the University of Milano Bicocca in Italy, Oxford Experimental Laboratory, Experimental Economics Laboratory-Royal Holloway in London, Finance and Economics Experimental Laboratory at Exeter, and ESSEXLab at Essex in Britain, Learning & Experimental Economics Projects at University of California-Santa Cruz, Social Science Experiments Lab at University of Colorado-Boulder, Appalachian Experimental Economics Laboratory in Boone, North Carolina, Center for Behavioral Political Economy in Stony Brook, New York, and University of Hawaii Laboratory for Computer-Mediated Experiments and the Study of Culture in Honolulu, Hawaii, in the US, and the Behavioural lab in Stockholm and Behavioural and Experimental Economics in Gothenburg in Sweden.

Table 2
Social value orientation allocation decisions.

	Allocation								
Decision	1	2	3	4	5	6	7	8	9
1	(85, 85)	(85, 76)	(85, 68)	(85, 59)	(85, 50)	(85, 41)	(85, 33)	(85, 24)	(85, 15)
2	(85, 15)	(87, 19)	(89, 24)	(91, 28)	(93, 33)	(94, 37)	(96, 41)	(98, 46)	(100, 50)
3	(50, 100)	(54, 98)	(59, 96)	(63, 94)	(68, 93)	(72, 91)	(76, 89)	(81, 87)	(85, 85)
4	(50, 100)	(54, 89)	(59, 79)	(63, 68)	(68, 58)	(72, 47)	(76, 36)	(81, 26)	(85, 15)
5	(100, 50)	(94, 56)	(88, 63)	(81, 69)	(75, 75)	(69, 81)	(63, 88)	(56, 94)	(50, 100)
6	(100, 50)	(98, 54)	(96, 59)	(94, 63)	(93, 68)	(91, 72)	(89, 76)	(87, 81)	(85, 85)

Notes: In each allocation, the first value is the number of tokens the decision-maker keeps for themselves and the second value is the number of tokens the other person receives.

Table 3
Multiple price list risk preference elicitation task.

Decision	Option A	Option B
1	15% chance of 100 tokens 85% chance of 40 tokens	70 tokens
2	25% chance of 100 tokens 75% chance of 40 tokens	70 tokens
3	35% chance of 100 tokens 65% chance of 40 tokens	70 tokens
4	45% chance of 100 tokens 55% chance of 40 tokens	70 tokens
5	55% chance of 100 tokens 45% chance of 40 tokens	70 tokens
6	65% chance of 100 tokens 35% chance of 40 tokens	70 tokens
7	75% chance of 100 tokens 25% chance of 40 tokens	70 tokens
8	85% chance of 100 tokens 15% chance of 40 tokens	70 tokens
9	95% chance of 100 tokens 5% chance of 40 tokens	70 tokens
10	100% chance of 100 tokens 0% chance of 40 tokens	70 tokens

3. Results

3.1. Analysis of compliance rates

We begin the analysis by reporting the average compliance rate for each value of the multiplier on the public good for each country and pooled across countries in Table 4. The table reveals the compliance rate increased significantly as the multiplier on the public good was increased. Overall, compliance increased by about 38% when the multiplier was increased from zero to two. This pattern of behavior is consistent across countries. Hence, the data appear to be consistent with the prediction in hypothesis 1.

Table 5 explores whether this pattern of behavior differs by gender. Columns 2 through 4 and 6 through 8 report female and male compliance rates, respectively, for each level of the public good multiplier across countries. Again, for both females and males we observe a significant increase in compliance as the public good multiplier increases.

Two additional patterns of behavior are evident in the table. First, in every single instance, the compliance rate for females is higher than the corresponding compliance rate for males, consistent with the behavior predicted in hypothesis 2. In particular, the female compliance rates are much higher when the public good is absent ($m = 0$). Second, there is a tendency for this gender gap in tax compliance to decrease as the public good multiplier increases, with the exception being the U.K. sample.[9] Overall, comparing compliance when the public good multiplier is zero ($m = 0$) to when it is two ($m = 2$), the gender gap decreases by roughly 42%; male compliance is 41% lower than females when the multiplier is zero but only 18% lower when the multiplier is two. This is consistent with hypothesis 3.

In order to formally test our hypotheses, we estimate the model in Eq. (1) using ordinary least squares regression analysis for each country separately, as well as the overall pooled sample. The results are presented in Table 6. [10] All models are estimated with subject-specific cluster-robust standard errors to account for the repeated observation of participants. The results strongly support hypothesis 1. In all countries, there is a significant increase in compliance when the public good multiplier is increased from zero to one. Moreover, in all four countries compliance is significantly higher when the public good multiplier is two than when it is one.[11] This is our first result.

Result 1: *There is a significant increase in compliance associated with increasing the public good multiplier, as predicted.*

We also find strong support for hypothesis 2. In all countries, compliance for females is significantly greater than for males. This is our second result.

Result 2: *The compliance rate of females is higher than that of males, as predicted.*

The results, however, are mixed regarding hypothesis 3. Although most interaction terms between females and the public good multipliers reported in the fifth and sixth rows of Table 6 are negative, they are mostly insignificant, with the exception of Italy, when the multiplier is one. However, we find evidence consistent with hypothesis 3 in Italy and Sweden when the multiplier is two.[12] These results are consistent with previous evidence that males contribute more than females to public goods. Controlling for pure tax compliance when the public good multiplier is zero, the propensity to contribute to a public good, captured by the increase in compliance, is greater for males than females when there is a significant difference. This is our third result.

Result 3: *When there is a significant difference between genders, the increase in compliance associated with an increase in the public good multiplier is greater for males than females, as predicted.*

3.2. Analysis of extensive and intensive margins of compliance

Next we investigate whether the differences in compliance rates across genders are due to differences in the fraction of participants underreporting their income and/or by the amount of income underreported (i.e., the intensive and/or extensive mar-

[9] The gender gap in tax compliance persists as the public good multiplier is increased in the U.K.

[10] Table 10 in the appendix reports the results of tobit analysis as robustness check, since the dependent variable is bounded between 0 and 1.

[11] In fact, compliance appears to be increasing linearly in the public good multiplier, m, given the estimated coefficient when $m = 2$ is roughly twice as large as the coefficient when $m = 1$.

[12] Men in the U.S. stand out because they appear to be more "purely" tax compliant than in other country samples. Moreover, since the gender gap is quite small in the U.S. with respect to the other samples, there is less room for it to improve.

Table 4
Compliance rate by treatment across countries.

Treatment				
Country	$m = 0$	$m = 1$	$m = 2$	Means test
Italy ($N = 415$)	0.596 (0.422)	0.660 (0.411)	0.784 (0.360)	49.11***
Sweden ($N = 327$)	0.515 (0.460)	0.640 (0.442)	0.799 (0.371)	73.15***
U.K. ($N = 360$)	0.347 (0.419)	0.430 (0.444)	0.596 (0.447)	58.80***
U.S. ($N = 537$)	0.604 (0.433)	0.691 (0.403)	0.773 (0.377)	50.27***
All ($N = 1639$)	0.528 (0.444)	0.616 (0.434)	0.742 (0.396)	224.46***

Notes: Average compliance rates for each country and treatment are reported with standard deviations in parentheses. The *F*-test for equal means using subject-specific cluster-robust standard errors is reported in the fifth column with statistical significance indicated by asterisks: *** indicates the difference is significant at the 1% level.

Table 5
Compliance rate by gender and treatment across countries.

Females Treatment					Males Treatment				
Country	$m = 0$	$m = 1$	$m = 2$	Means test	Country	$m = 0$	$m = 1$	$m = 2$	means test
Italy ($N = 146$)	0.697 (0.393)	0.724 (0.395)	0.815 (0.348)	10.16***	Italy ($N = 164$)	0.389 (0.427)	0.527 (0.451)	0.712 (0.408)	42.51***
Sweden ($N = 143$)	0.699 (0.423)	0.807 (0.355)	0.883 (0.282)	17.84***	Sweden ($N = 184$)	0.372 (0.438)	0.511 (0.459)	0.734 (0.417)	59.56***
U.K. ($N = 156$)	0.482 (0.443)	0.557 (0.419)	0.731 (0.388)	30.50***	U.K. ($N = 200$)	0.234 (0.363)	0.325 (0.435)	0.487 (0.463)	29.26***
U.S. ($N = 298$)	0.661 (0.413)	0.758 (0.357)	0.808 (0.347)	25.04***	U.S. ($N = 223$)	0.507 (0.447)	0.592 (0.441)	0.714 (0.414)	26.51***
All (N = 743)	0.638 (0.425)	0.719 (0.387)	0.808 (0.348)	80.43***	All ($N = 771$)	0.379 (0.431)	0.490 (0.457)	0.659 (0.438)	148.70***

Notes: Average compliance rates for each country and treatment are reported with standard deviations in parentheses. The *F*-test for equal means using subject-specific cluster-robust standard errors is reported in the fifth column with statistical significance indicated by asterisks: *** indicates the difference is significant at the 1% level.

Table 6
Regression analysis of compliance.

Variable	Italy	Sweden	U.K.	U.S.	Overall
Constant ($m = 0$)	0.389***	0.372***	0.234***	0.507***	0.379***
	(0.033)	(0.032)	(0.026)	(0.030)	(0.016)
Single ($m = 1$)	0.138***	0.139***	0.091***	0.085***	0.111***
	(0.034)	(0.030)	(0.025)	(0.028)	(0.014)
Double ($m = 2$)	0.324***	0.362***	0.254***	0.207***	0.281***
	(0.036)	(0.034)	(0.033)	(0.030)	(0.017)
Female	0.308***	0.327***	0.248***	0.154***	0.259***
	(0.047)	(0.048)	(0.044)	(0.038)	(0.022)
Single∗Female	−0.111**	−0.030	−0.015	0.011	−0.030
	(0.044)	(0.042)	(0.038)	(0.034)	(0.019)
Double∗Female	−0.206***	−0.178***	−0.004	−0.060	−0.111***
	(0.047)	(0.046)	(0.047)	(0.037)	(0.022)
R-squared	0.112	0.161	0.126	0.058	0.105
Num. Obs.	930	981	1068	1563	4542

Notes: *, **, and *** denotes estimates that are statistically different from zero at the 10%, 5%, and 1% significance levels, respectively. Subject-specific cluster-robust standard errors are reported in parentheses.

gins). Table 7 summarizes the extensive margin of compliance by gender. The table reports the proportion of participants that report all of their income as percentages. There are a few behavioral patterns that are noteworthy. First, in every country and treatment the proportion of males that report all of their income is smaller than the corresponding proportion of females, which is consistent with hypothesis 2. Second, in every single country the extensive margin of compliance increases as the public good multiplier is increased for both genders. Finally, in every country except the U.K. the gender gap in the extensive margin of compliance declines as the public good multiplier is increased, which is consistent with hypothesis 3. These patterns in the extensive margin of compliance are consistent with the compliance rates reported in Table 5.

Turning our attention to the intensive margin of compliance, we examine the cumulative distributions of compliance for each gender conditional on underreporting income. Fig. 1 plots the distributions of non-compliance for each gender and country pooling the data across public good multipliers. The patterns of behavior are quite consistent across countries. In every instance, among participants that underreport income, the proportion that reports zero income is greater for males than females. This is the principle reason for the difference in the distributions, as the cumulative distributions do not tend to converge until compliance rates are fairly high. Hence, when underreporting income, males underreport by larger amounts; the intensive margin of compliance is lower for males, consistent with hypothesis 2.

Fig. 2 plots the distributions of non-compliance for each gender and public good multiplier pooling the data across countries. The graphs reveal an interesting behavioral asymmetry. While increasing the public good multiplier reduces the proportion of females

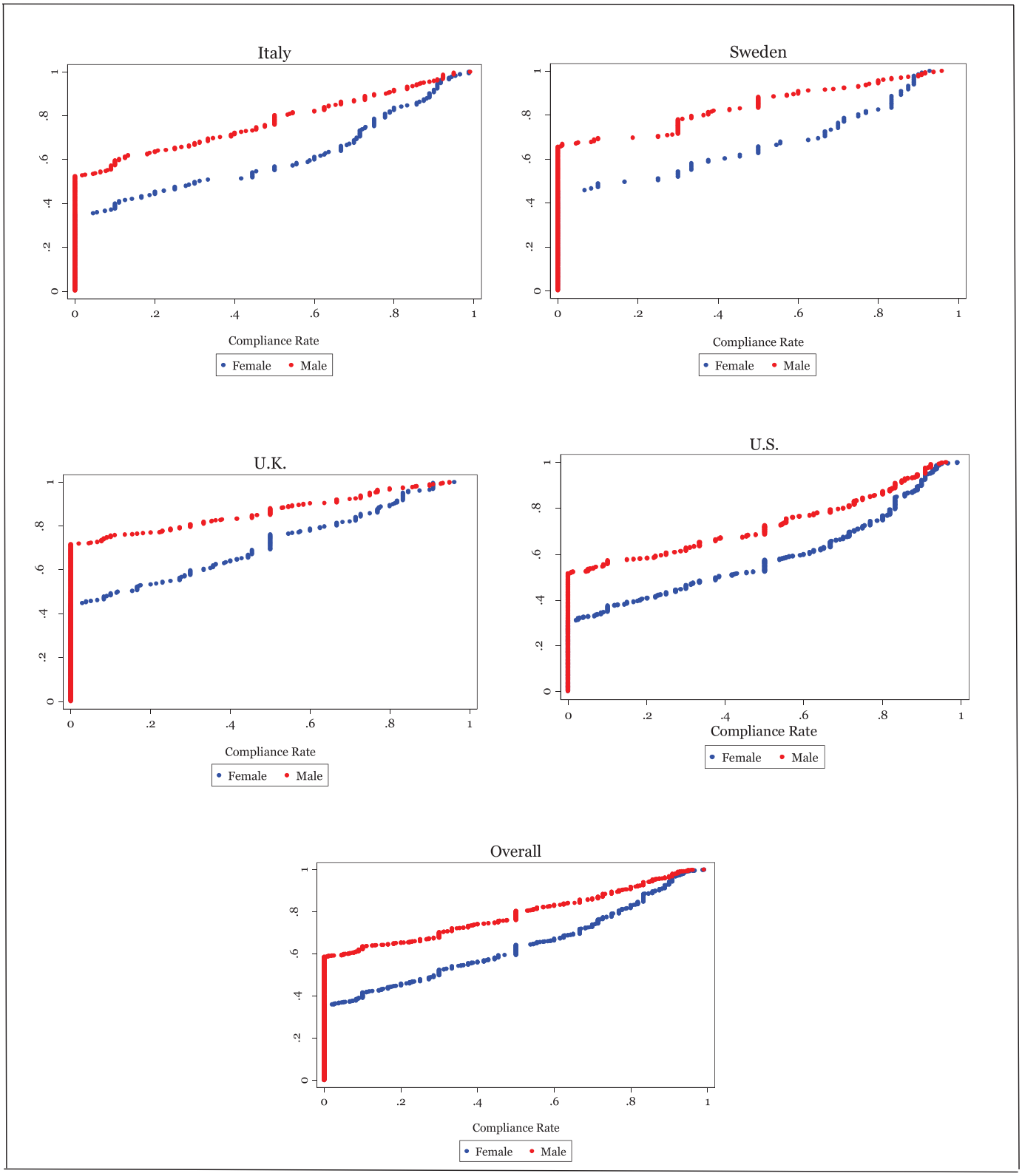

Fig. 1. Cumulative distribution of non-compliance by country and gender.

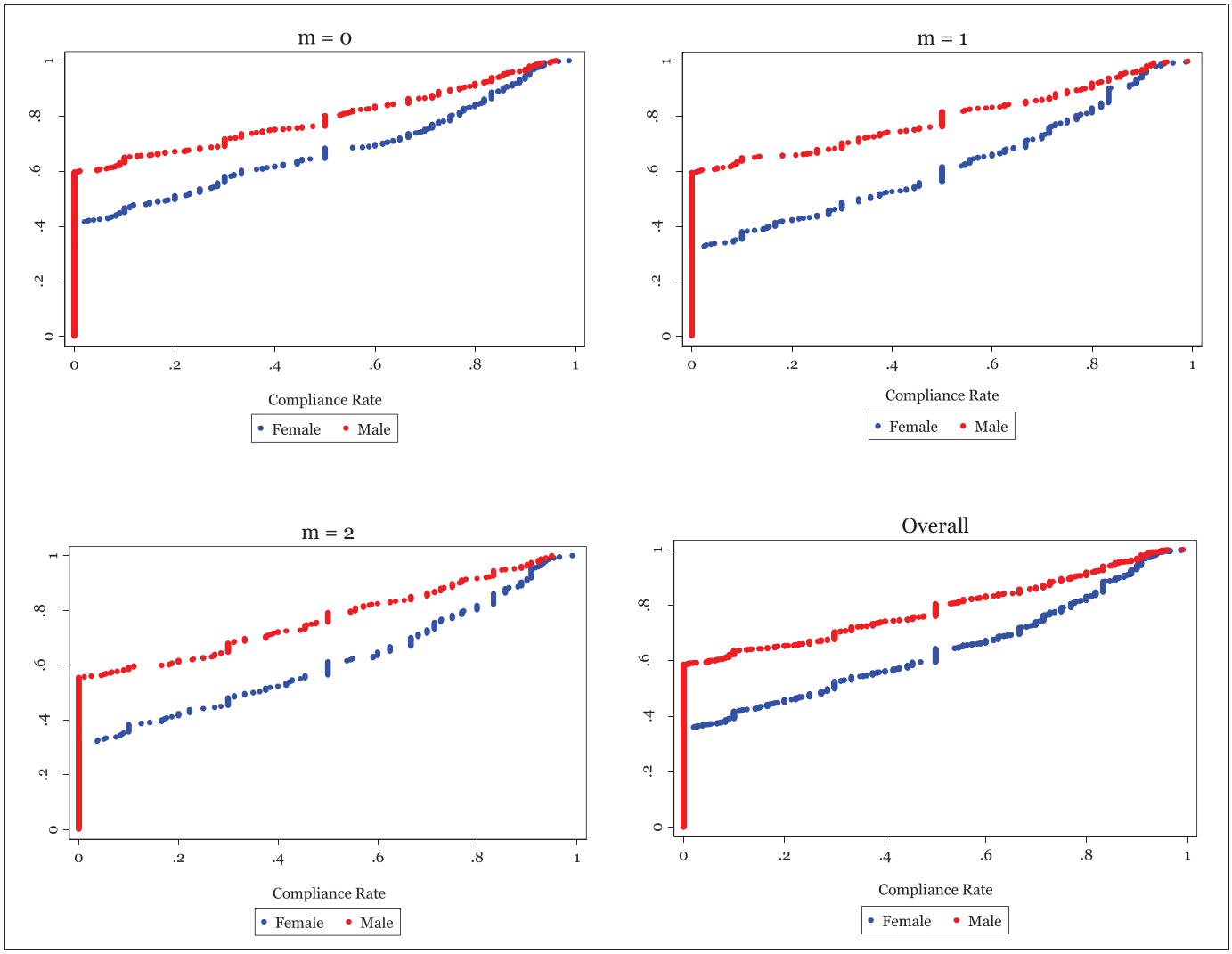

Fig. 2. Cumulative distribution of non-compliance by treatment and gender.

reporting zero income, we observe little effect for males. Hence, the reduction of the gender gap in tax compliance associated with increasing the public good multiplier can be attributed to changes in both the extensive and intensive margins across genders.

3.3. Robustness analysis

In this section we explore the robustness of the main results. In particular, we examine whether other demographic variables such as risk attitude, social preference, political orientation, and attitudes toward government can account for the gender gap in tax compliance and propensity to contribute public goods. Furthermore, we examine whether the main results are robust to variation in the treatment order.

As previously stated, the experiment elicited various demographic variables in addition to gender to be used as control variables in the analysis. Participants' social preferences were elicited using the incentive compatible allocation mechanism described in Table 3. Participants preferred allocation choice for the six decisions results in a continuous scale measure of their pro-sociality. This measure is standardized for the subsequent analysis. Moreover, the survey administered upon completion of the experiment asked subjects to indicate their risk attitude on a 10-point Likert scale. In addition, a series of questions elicited participants' political orientation, attitudes towards various government institutions, and attitudes towards tax compliance. Pampel et al. (2016) conduct a factor analysis of participants' survey responses which resulted in three determinants of tax compliance: (i) pro-welfare ideology (ii) duty to pay and (iii) trust in government. The results from including these additional demographic controls are shown in Table 8.[13]

The results of estimates of the model in Eq. (1) using ordinary least squares regression analysis for each country separately, as well as the overall combined sample. Again, all models are estimated with subject-specific cluster-robust standard errors to account for the repeated observation of participants. The main results concerning the effects of treatment dummies, the gender dummy, and the interactions of treatments and gender are fairly consistent across countries. Moreover, these results are quite consistent with the previous results reported in Table 6. The main results appear to be robust to the inclusion of additional demographic controls.

Nonetheless, the additional demographic controls are significantly correlated with tax compliance. Social and risk preferences are both consistently and significantly correlated with tax com-

[13] Table 11 in the appendix reports the results of tobit analysis as robustness check, since the dependent variable is bounded between 0 and 1.

Table 7
Extensive margins of compliance.

Females Treatment				Males Treatment			
Country	$m = 0$	$m = 1$	$m = 2$	Country	$m = 0$	$m = 1$	$m = 2$
Italy ($N = 146$)	47.9	56.2	70.5	Italy ($N = 164$)	20.7	40.9	60.4
Sweden ($N = 143$)	58.0	70.6	79.7	Sweden ($N = 184$)	26.6	40.8	67.4
U.K. ($N = 156$)	32.7	37.2	58.3	U.K. ($N = 200$)	11.5	24.0	38.5
U.S. ($N = 298$)	46.3	54.7	67.4	U.S. ($N = 223$)	33.2	42.2	61.9
All ($N = 743$)	46.0	54.4	68.5	All ($N = 771$)	23.3	36.8	56.8

Table 8
Regression analysis of compliance with demographic controls.

Variable	Italy	Sweden	U.K.	U.S.	Overall
Constant ($m = 0$)	0.419***	0.391***	0.686***	0.626***	0.581***
	(0.082)	(0.093)	(0.109)	(0.083)	(0.041)
Single ($m = 1$)	0.133***	0.139***	0.091***	0.085***	0.102***
	(0.035)	(0.030)	(0.025)	(0.028)	(0.013)
Double ($m = 2$)	0.321***	0.362***	0.255***	0.207***	0.260***
	(0.037)	(0.034)	(0.034)	(0.030)	(0.016)
Female	0.253***	0.274***	0.166***	0.112***	0.174***
	(0.047)	(0.046)	(0.044)	(0.037)	(0.020)
Single∗Female	−0.102**	−0.030	−0.015	0.011	−0.023
	(0.046)	(0.043)	(0.038)	(0.034)	(0.018)
Double∗Female	−0.198***	−0.178***	−0.004	−0.060	−0.095***
	(0.048)	(0.046)	(0.047)	(0.037)	(0.020)
Social preference	0.005***	0.009***	0.007***	0.006***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Risk preference	−0.022***	−0.022***	−0.033***	−0.022***	−0.025***
	(0.008)	(0.008)	(0.009)	(0.007)	(0.004)
Income	0.000	−0.001	−0.002***	−0.001*	−0.001***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Employment	−0.012	−0.004	−0.017	−0.002	−0.003
	(0.043)	(0.033)	(0.037)	(0.028)	(0.016)
Experience	−0.055	−0.080**	−0.125***	−0.079***	−0.104***
	(0.049)	(0.037)	(0.046)	(0.028)	(0.016)
Welfare ideology	0.045	0.096***	0.040	0.027	0.068***
	(0.036)	(0.032)	(0.032)	(0.028)	(0.014)
Duty to pay	0.119***	0.079**	0.089**	0.036	0.065***
	(0.036)	(0.035)	(0.037)	(0.028)	(0.015)
Government	−0.046	−0.050	−0.054	−0.071**	−0.064***
Trust	(0.035)	(0.038)	(0.037)	(0.031)	(0.015)
R-squared	0.220	0.317	0.266	0.148	0.228
Num. Obs.	882	981	1062	1563	4488

Notes: *, **, and *** denotes estimates that are statistically different from zero at the 10%, 5%, and 1% significance levels, respectively. Subject-specific cluster-robust standard errors are reported in parentheses.

pliance in the anticipated directions. Pro-social preferences, risk aversion, and obligatory attitudes towards tax payments are all positively correlated with tax compliance.[14] Trust in government is significantly correlated with compliance in the U.K. and the U.S. Additionally, there is some evidence that favorable attitudes towards the welfare state are positively correlated with compliance in Sweden.

Finally, the robustness of the results to variation in the treatment order is investigated. As previously stated, six additional sessions were conducted in Italy using treatment order B in Table 9.[15] In addition to varying the treatment order, these sessions included an incentive compatible risk preference elicitation mechanism shown in Table 3 to further verify that risk attitudes are not a determinant of the gender gap in tax compliance. Three models are estimated using ordinary least squares with subject-specific cluster-robust standard errors to account for the repeated observation of participants. The first model estimates simple treatment effects, the second allows the treatment effects to vary by gender, and the third allows the treatment effects to vary by gender while controlling for additional demographics.[16]

In general, the results in Table 9 lend further support to the main results. All the models indicate there is an increase in compliance when the public good multiplier is increased from zero, although the effect is only significant in the first model. All the models indicate female compliance is significantly greater than that of males, even when controlling for risk preference. Hence, gender gap in tax compliance is particularly robust. Unlike the results from treatment order A, there is no significant evidence that males are more sensitive to the return on the public good in treatment order B.

[14] The negative coefficients on risk preference indicate that as risk aversion declines so does compliance.

[15] Table 12 in the appendix reports the results of tobit analysis as robustness check, since the dependent variable is bounded between 0 and 1.

[16] We allowed these demographic effects to vary by gender. Only social preference had a significant interaction with gender, the remainder was jointly insignificant and hence dropped from the model.

Table 9
Regression analysis of compliance for treatment order B.

Variable	Model 1	Model 2	Model 3
Constant ($m = 0$)	0.539***	0.549***	0.071
	(0.053)	(0.057)	(0.151)
Single ($m = 1$)	0.082***	0.076	0.080
	(0.031)	(0.052)	(0.061)
Double ($m = 2$)	0.067**	0.041	0.036
	(0.033)	(0.052)	(0.060)
Female	0.216***	0.196***	0.458***
	(0.062)	(0.071)	(0.133)
Single∗Female		0.011	0.022
		(0.063)	(0.073)
Double∗Female		0.047	0.066
		(0.067)	(0.078)
Social preference			0.014***
			(0.004)
Social preference∗			−0.012**
Female			(0.005)
Risk preference			0.030*
			(0.016)
Welfare ideology			−0.020
			(0.060)
Duty to pay			−0.036
			(0.045)
Government trust			−0.070
			(0.060)
R-squared	0.084	0.085	0.220
Num. Obs.	360	360	300

Notes: *, **, and *** denotes estimates that are statistically different from zero at the 10%, 5%, and 1% significance levels, respectively. Subject-specific cluster-robust standard errors are reported in parentheses.

4. Discussion

We began with the question, are there gender differences in willingness to pay for public goods through taxation? There are reasons to think they should exist, but predicting the direction in which the genders differ is problematic. On one hand, there is evidence that women are more compliant when it comes to reporting their income for tax purposes. On the other hand, there is also evidence that women may be more willing to free-ride when it comes to voluntarily contributing to the provision of public goods. Given the potential for these two effects to offset each other, it remains an open empirical question as to which dominates when public goods are financed through taxation.

To investigate the issue, we utilize data from laboratory experiments conducted in Italy, Sweden, U.K. and the U.S. Subjects earned income in a real effort task and reported their income for tax purposes, which were either used to finance a public good or not. Consequently, we are able to separately identify tax compliance from willingness to contribute to the public good. The size and scope of this study makes it the most comprehensive investigation to date. Furthermore, conducting experiments in multiple countries allows us to test for cultural differences in gender effects.

The results suggest that women pay more for public goods, although men may be more sensitive to the price of provision. Of course, this conclusion is constrained to the parameter values used in the experiment; it could be true than men pay more for larger marginal per capita returns from the public good. Nonetheless, we find robust evidence across countries in the sample that women are more compliant when paying taxes. On the other hand, there is only significant difference across genders in the willingness to contribute to public goods in Italy and Sweden; suggesting culture may be an important factor. The data analysis reveals the gender gap in compliance is due to differences in both the extensive and intensive margins; men are more likely to underreport their income and when they do, they tend to evade all of their tax liability. Finally, contrary to conventional wisdom, a robustness analysis indicates the gender gap in compliance is not due to differences in risk aversion.

Of course, the extent to which these results can be applied beyond the laboratory depends on the degree of "parallelism" to the naturally occurring world (Smith, 1982; Plott, 1987). The experimental setting need not attempt to capture all of the variation in the naturally occurring environment, but it should sufficiently recreate the fundamental elements if the results are to be relevant in policy debates. Our experimental design implements the fundamental elements of any voluntary tax reporting system. Moreover, there is evidence of the external validity of this decision setting (Alm et al., 2015). Hence, we feel confident that the reported behavior is likely to occur outside of the lab as well.

The robustness of the gender gap in tax compliance has important policy implications. Modern societies continue to be plagued by tax evasion and reducing the tax gap continues to be a principle object of tax revenue agencies in all countries. As welfare states adapt to a challenging set of demographic, economic, and fiscal pressures, a state's ability to extract revenue from its citizens is crucial to sustaining a well-functioning welfare regime. While policies are typically focused on improving the enforcement regime, this research highlights the potential for a different avenue for policy makers. By improving labor market outcomes such as wage gaps and labor force participation rates for women, governments can potentially make significant reductions in tax gaps, given tax compliance is higher for women.

Acknowledgments

We would like to thank Giulia Andrighetto, Sanne Noyon, Stefania Ottone, Ferruccio Ponzano, Nan Zhang, and the lab personnel at Bologna Laboratory for Experiments in Social Sciences, Centro d'Economia Sperimentale A Roma Est, and Experimental Economics Lab of the University of Milano Bicocca in Italy, Oxford Experimental Laboratory, Experimental Economics Laboratory-Royal Holloway in London, Finance and Economics Experimental Laboratory at Exeter, and ESSEXLab at Essex in Britain, Learning & Experimental Economics Projects at University of California-Santa Cruz, Social Science Experiments Lab at University of Colorado-Boulder, Appalachian Experimental Economics Laboratory in Boone, North Carolina, Center for Behavioral Political Economy in Stony Brook, New York, and University of Hawaii Laboratory for Computer-Mediated Experiments and the Study of Culture in Honolulu, Hawaii, in the US, and the Behavioural lab in Stockholm and Behavioural and Experimental Economics in Gothenburg in Sweden for all of their assistance. All errors or omissions are our responsibility. Funds for this research were provided by the European Research Council (Grant agreement number 295675).

Appendix

Tables 10–12.

Table 10
Tobit regression analysis of compliance.

Variable	Italy	Sweden	U.K.	U.S.	Overall
Constant ($m = 0$)	0.123	−0.058	−0.482***	0.487***	0.058
	(0.107)	(0.159)	(0.130)	(0.089)	(0.058)
Single ($m = 1$)	0.499***	0.625***	0.315***	0.262***	0.401***
	(0.116)	(0.151)	(0.105)	(0.086)	(0.054)
Double ($m = 2$)	1.131***	1.740***	0.929***	0.734***	1.049***
	(0.150)	(0.224)	(0.146)	(0.108)	(0.072)
Female	0.929***	1.439***	0.922***	0.473***	0.885***
	(0.158)	(0.257)	(0.175)	(0.118)	(0.082)
Single∗Female	−0.347**	−0.001	−0.047	0.036	−0.095
	(0.152)	(0.229)	(0.137)	(0.109)	(0.071)
Double∗Female	−0.539***	−0.573**	−0.040	−0.169	−0.312***
	(0.178)	(0.262)	(0.172)	(0.128)	(0.085)
Psuedo *R*-squared	0.058	0.082	0.061	0.031	0.052
Num. Obs.	930	981	1068	1563	4542

Notes: *, **, and *** denotes estimates that are statistically different from zero at the 10%, 5%, and 1% significance levels, respectively. Subject-specific cluster-robust standard errors are reported in parentheses.

Table 11
Tobit analysis of compliance with demographic controls.

Variable	Italy	Sweden	U.K.	U.S.	Overall
Constant ($m = 0$)	1.100***	1.519***	1.198***	1.088***	1.210***
	(0.082)	(0.141)	(0.084)	(0.061)	(0.042)
Single ($m = 1$)	0.483***	0.609***	0.320***	0.268***	0.394***
	(0.115)	(0.148)	(0.104)	(0.086)	(0.053)
Double ($m = 2$)	1.116***	1.712***	0.924***	0.732***	1.040***
	(0.151)	(0.222)	(0.147)	(0.108)	(0.072)
Female	0.719***	1.089***	0.603***	0.334***	0.624***
	(0.152)	(0.227)	(0.159)	(0.110)	(0.075)
Single∗Female	−0.323**	0.028	−0.047	0.022	−0.087
	(0.151)	(0.223)	(0.136)	(0.108)	(0.071)
Double∗Female	−0.524***	−0.554**	−0.055	−0.178	−0.318***
	(0.180)	(0.255)	(0.169)	(0.127)	(0.085)
Social preference	0.016***	0.043***	0.025***	0.019***	0.025***
	(0.004)	(0.008)	(0.005)	(0.004)	(0.002)
Risk preference	−0.081***	−0.100**	−0.123***	−0.073***	−0.095***
	(0.028)	(0.042)	(0.034)	(0.022)	(0.015)
Income	0.001	−0.004	−0.007***	−0.003**	−0.004***
	(0.001)	(0.003)	(0.002)	(0.001)	(0.001)
Employment	−0.067	−0.046	−0.021	0.003	0.018
	(0.139)	(0.170)	(0.133)	(0.094)	(0.062)
Experience	−0.199	−0.397**	−0.437**	−0.272***	−0.348***
	(0.161)	(0.199)	(0.173)	(0.094)	(0.069)
Welfare ideology	0.154	0.483***	0.151	0.098	0.260***
	(0.124)	(0.161)	(0.126)	(0.093)	(0.058)
Duty to pay	0.360***	0.434**	0.309**	0.186**	0.316***
	(0.119)	(0.180)	(0.141)	(0.092)	(0.061)
Government	−0.141	−0.218	−0.199	−0.221**	−0.229***
Trust	(0.120)	(0.203)	(0.140)	(0.103)	(0.062)
Psuedo *R*-squared	0.116	0.182	0.146	0.082	0.131
Num. Obs.	882	981	1062	1563	4488

Notes: *, **, and *** denotes estimates that are statistically different from zero at the 10%, 5%, and 1% significance levels, respectively. Subject-specific cluster-robust standard errors are reported in parentheses.

Table 12
Tobit regression analysis of compliance for treatment order B.

Variable	Model 1	Model 2	Model 3
Constant ($m = 0$)	0.674***	0.773***	−0.424
	(0.112)	(0.113)	(0.342)
Single ($m = 1$)	0.139*	0.008	0.012
	(0.084)	(0.121)	(0.138)
Double ($m = 2$)	0.114	−0.065	−0.082
	(0.086)	(0.127)	(0.144)
Female	0.524***	0.333**	0.839***
	(0.151)	(0.144)	(0.276)
Single∗Female		0.251	0.317*
		(0.171)	(0.187)
Double∗Female		0.348**	0.420**
		(0.176)	(0.193)
Social preference			0.028***
			(0.009)
Social preference∗			−0.028**
Female			(0.011)
Risk preference			0.094**
			(0.042)
Welfare ideology			−0.026
			(0.146)
Duty to pay			−0.092
			(0.097)
Government trust			−0.178
			(0.128)
Pseudo R-squared	0.039	0.042	0.111
Num. Obs.	360	360	300

Notes: *, **, and *** denotes estimates that are statistically different from zero at the 10%, 5%, and 1% significance levels, respectively. Subject-specific cluster-robust standard errors are reported in parentheses.

References

Alm, J., Bloomquist, K.M., McKee, M., 2015. On the external validity of laboratory tax compliance experiments. Econ. Inq. 53 (2), 1170–1186.
Alm, J., Cherry, T., Jones, M., McKee, M., 2010. Taxpayer information assistance services and tax compliance behavior. J. Econ. Psychol. 31 (4), 577–586.
Alm, J., Cherry, T.L., Jones, M., McKee, M., 2012. Social programs as positive inducements for tax participation. J. Econ. Behav. Organ. 84 (1), 85–96.
Alm, J., Jackson, B.R., McKee, M., 2009. Getting the word out: Enforcement information dissemination and compliance behavior. J. Public Econ. 93 (3), 392–402.
Alm, J., McClelland, G.H., Schultz, W.D., 1992. Why do people pay taxes? J. Public Econ. 48, 21–38.
Alm, J., McKee, M., 1998. Extending the lessons of laboratory experiments on tax compliance to managerial and decision economics. Managerial and Decision Economics 259–275.
Alm, J., Torgler, B., 2006. Culture differences and tax morale in the United States and in Europe. J. Econ. Psychol. 27 (2), 224–246.
Andersen, S.G., Harrison, G., Lau, M.I., Rutström, E.E., 2006. Elicitation using multiple price list formats. Exp. Econ. 9 (4), 383–406.
Andreoni, J., Vesterlund, L., 2001. Which is the fair sex? Gender differences in altruism. Q. J. Econ. 116 (1), 293–312.
Brown-Kruse, J., Hummels, D., 1993. Gender effects in laboratory public goods contribution: do individuals put their money where their mouth is? J. Econ. Behav. Organ. 22 (3), 255–267.
Bucciol, A., Piovesan, M., 2011. Luck or cheating? A field experiment on honesty with children. J. Econ. Psychol. 32 (1), 73–78.
Cadsby, C.B., Maynes, E., 1998. Gender and free riding in a threshold public goods game: experimental evidence. J. Econ. Behav. Organ. (34) 603–620.
Cadsby, C.B., Maynes, E., Trivedi, V.U., 2006. Tax compliance and obedience to authority at home and in the lab: a new experimental approach. Exp. Econ. 9 (4), 343–359.
Cappelen, A.W., Sørensen, E.Ø., Tungodden, B., 2013. When do we lie? J. Econ. Behav. Organ. 93, 258–265.
Charness, G., Gneezy, U., 2012. Strong evidence for gender differences in risk taking. J. Econ. Behav. Organ. 83, 50–58.
Childs, J., 2012. Gender differences in lying. Econ. Lett. 114, 147–149.
Cox, J.C., Deck, C.A., 2006. When are women more generous than men? Econ. Inq. 44 (4), 587–598.
Croson, R., Gneezy, U., 2009. Gender Differences in Preferences. J. Econ. Liter. 47, 448–474.
Dreber, A., Johannesson, M., 2008. Gender differences in deception. Econ. Lett. 99 (1), 197–199.
Eckel, C.C., Grossman, P.J., 2008. Men, women and risk aversion: experimental evidence. In: Handbook of Experimental Economics Results, 1, pp. 1061–1073.
Eckel, C.C., Füllbrunn, S.C., 2015. Thar she blows? Gender, competition, and bubbles in experimental asset markets. Am. Econ. Rev. 105 (2), 906–920.
Fischbacher, U., 2007. Z-Tree – Zurich toolbox for readymade economic experiments – experimenter's manual. In: Experimental Economics, pp. 171–178.
Filippin, A., Crosetto, P., 2016. A reconsideration of gender differences in risk attitudes. Manag. Sci. 62 (11), 3138–3160.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125.
Gylfason, H.F., Arnardottir, A.A., Kristinsson, K., 2013. More on gender differences in lying. Econ. Lett. 119 (1), 94–96.
Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. Am. Econ. Rev. 92 (5), 1644–1657.
Isaac, R.M., McCue, K.F., Plott, C.R., 1985. Public goods provision in an experimental market. J. Public Econ. 26, 51–74.
Karakostas, A., Zizzo, D.J., 2016. Compliance and the power of authority. J. Econ. Behav. Organ. 124, 67–80.
Murphy, R.O., Ackermann, K.A., Handgraaf, M.J., 2011. Measuring Social Value Orientation. Judgment and Decision Making 6 (8), 771–781.
Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? Q. J. Econ. 122 (3), 1067–1101.
Niederle, M., Vesterlund, L., 2008. Gender differences in competition. Negotiation Journal 24 (4), 447–463.
Niederle, M., Vesterlund, L., 2011. Gender and competition. Annu. Rev. Econ. 3 (1), 601–630.
Nowell, C., Tinkler, S., 1994. The influence of gender on the provision of a public good. J. Econ. Behav. Organ. 25 (1), 25–36.
Pampel, F., G. Angrighetto, and S. Steinmo, (2016) Culture, institutions, and tax compliance: a cross-national experiment and survey. Working Paper.
Plott, C.R., 1987. Dimensions of parallelism: some policy applications of experimental methods. In: Roth, Alvin E. (Ed.), Laboratory Experimentation in Economics: Six Points of View. Cambridge University Press, New York, NY.
Smith, V.L., 1982. Microeconomic systems as an experimental science. Am. Econ. Rev. 72, 923–955.
Solow, J.L., Kirkwood, N., 2002. Group identity and gender in public goods experiments. J. Econ. Behav. Organ. 48 (4), 403–412.
Spicer, M.W., Becker, L.A., 1980. Fiscal inequity and tax evasion: An experimental approach. National Tax Journal 171–175.
Stockard, J., van de Kragt, A.J.C., Dodge, P.J., 1988. Gender roles and behavior in social dilemmas: are there sex differences in cooperation and in its justification? Soc. Psychol. Q. 51 (2), 154–163.
Torgler, B., 2002. Speaking to theorists and searching for facts: tax morale and tax compliance in experiments. J. Econ. Surv. 16 (5), 657–683.