NBER WORKING PAPER SERIES

WHAT MAKES A TAX EVADER?

Marcelo L. Bergolo
Martin Leites
Ricardo Perez-Truglia
Matias Strehl

Working Paper 28235
http://www.nber.org/papers/w28235

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2020, Revised December 2021

We are thankful for excellent comments from colleagues and seminar discussants. We thank the Dirección General Impositiva de Uruguay and Centro de Estudios Fiscales for their support. This research project received the financial support from the Laboratorio de Políticas Públicas (OPP, Uruguay). The survey was pre-registered in the AEA RCT Registry (#0004108). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

What Makes a Tax Evader?
Marcelo L. Bergolo, Martin Leites, Ricardo Perez-Truglia, and Matias Strehl
NBER Working Paper No. 28235
December 2020, Revised December 2021
JEL No. H24,H26,K42,Z1,Z13

ABSTRACT

Why do some individuals choose to evade taxes while others do not? One popular view is that some individuals cheat on their taxes because they are more dishonest, selfish, or perceive different social norms. There is, however, little direct evidence on this matter. In collaboration with the national tax agency in Uruguay, we address this question using a combination of surveys and administrative records. Leveraging a unique institutional setting, we measure individual-level evasion choices. We document significant variation in evasion decisions across individuals. For a subsample of 6,078 taxpayers, we use survey questions and incentivized laboratory games to measure traits such as honesty, selfishness, and perceived social norms. We find that these individual characteristics have some power to predict who evades taxes. However, other factors, such as the marginal tax rates and the behavior of peers, play a bigger role.

Marcelo L. Bergolo
Instituto de Economia (IECON)
Universidad de La Republica
1926 Gonzalo Ramirez
Montevideo
Uruguay
and IZA
marcelo.bergolo@fcea.edu.uy

Martin Leites
Instituto de Economia (IECON)
Universidad de La Republica
1926 Gonzalo Ramirez
Montevideo
Uruguay
mleites@iecon.ccee.edu.uy

Ricardo Perez-Truglia
Haas School of Business
University of California, Berkeley
545 Student Services Building #1900
Berkeley, CA 94720-1900
and NBER
ricardotruglia@berkeley.edu

Matias Strehl
Instituto de Economia (IECON)
Universidad de La Republica
1926 Gonzalo Ramirez
Montevideo, Uruguay
Uruguay
matias.strehl@gmail.com

A data appendix is available at http://www.nber.org/data-appendix/w28235

1 Introduction

Some individuals evade taxes, whereas others do not. What drives these differences in behavior? One popular view attributes these differences in behavior to differences in tax morale. More precisely, the hypothesis is that some traits, such as dishonesty or selfishness, or their beliefs, such as perceived social norms, make some people more likely to evade taxes. There is, however, no direct evidence on this matter. The lack of evidence is perhaps not surprising given the data challenges. Measuring evasion choices at the individual level is difficult on its own (Slemrod, 2019). In addition, testing this hypothesis requires individual-level data on the values and beliefs of the taxpayers. We collaborated with Uruguay's national tax agency and collected all the necessary data to address this question.

The first part of the study presents an individual-level measure of tax evasion from a natural, high-stakes context. Measuring tax evasion is challenging, especially at the individual level. With rare exception, the government cannot observe what individuals *should* have declared on their tax forms; thus, researchers cannot observe it either (Slemrod, 2019).[1] We exploit a unique feature of the institutional setting to observe evasion decisions at the individual level.

Our main measure of tax evasion is based on the under-reporting of wages. Taxpayers pay income taxes for each additional dollar that they report in wages and therefore have incentive to under-report wages to reduce their tax liability. As in most of the world, individual taxpayers in Uruguay must file an annual tax return to declare their incomes, deductions, and tax withholdings. Uruguay also uses a third-party reporting system wherein employers send information (e.g., wages paid) directly to the tax authority. In some countries, like the United States, the tax agency automatically corrects any discrepancies between the individual tax return and the third-party report.[2] In developing countries, however, there is more limited enforcement of third-party reporting (Brockmeyer and Hernandez, 2016; Carrillo et al., 2017). As a result, even when the information is subject to third-party reporting, taxpayers can still evade taxes by under-reporting their wages in the tax returns they file. We compare the information reported by the taxpayers with the information reported by their employers and use the gaps to identify which taxpayers under-report their wages and by how much (Best, 2014; Bergolo et al., 2019).

[1] Some exceptions include the subset of individuals who are audited each year, from which one can observe the evasion amounts that were detected in the audit. However, only a small, non-random minority of individuals are audited, so it is possible that these audits correspondingly uncover only a small and non-random share of the amount evaded.

[2] For example, if a U.S. employee under-reports the wage relative to the employer's third-party report, the Internal Revenue Service automatically corrects the individual's tax form, updates the tax amount, and notifies the taxpayer of the correction.

We study all 151,565 taxpayers in the country who earned wages and filed a tax return in 2016. We find that 15.5% of them under-reported their wages and, among those who under-reported wages, on average evaded about 17% of their tax liability (equivalent to U$D 344). We provide evidence that these discrepancies are not due to random mistakes. For example, we show that income under-reporting is much more likely than over-reporting: whereas 15.5% of taxpayers under-report wages, only 3.9% over-report their wages. The disproportionate amount of mistakes in the direction that best serves the taxpayers' financial interest suggests that wage under-reporting is largely intentional.

Albeit suggestive, this evidence does not prove that the misreporting corresponds to tax evasion. For example, it is possible that employees make random mistakes and that downward mistakes are more likely than upward mistakes. We leverage a unique feature of the institutional context to rule out this confound. In addition to the incentive to misreport their wages, employees have incentives to misreport two other items in the tax form: tax deductions and tax withholdings. However, while employees must *under-report* wages to lower their tax liability, they must *over-report* their deductions and withholdings if they want to evade taxes. We find systematic over-reporting of deductions and withholdings. The fraction of individuals who over-report deductions (15.8%) is substantially higher than the fraction who under-report (2.2%). Likewise, the share of over-reporters of withholdings (22.7%) is substantially larger than the share of under-reporters (2.9%).[3] The fact that taxpayers under-report wages but over-report deductions and withholdings indicates that the misreporting is largely due to intentional tax evasion and not just random mistakes.

While the margin of tax evasion that we study is not the main source of tax evasion in the country, it is arguably best suited for studying differences in evasion choices across individuals. For example, business owners can under-report revenues or over-report costs to evade value added taxes (VAT). However, the decision to evade VAT may be largely related to the strength of tax enforcement in the industry. For example, retailers may have an easier time under-reporting revenue than upstream businesses (Pomeranz, 2015; Naritomi, 2019). As a result, the choice to evade VAT may say more about the strength of the paper trail in that industry than about the preferences of the business owner. In contrast, the strength of third-party reporting is arguably the same for all employees in the country, and thus wage under-reporting is best suited for inter-personal comparisons.

One useful feature of our data is that we observe multiple decisions from the same taxpayer. In a given year, a taxpayer can choose to under-report wages and then face the same opportunity a year later. We show that individuals who under-report wages one year are

[3] For simplicity, these results on deduction and withholding misreporting are based on the sample of taxpayers who report their wages truthfully. See Section 3 for more details.

more likely to under-report wages on the next year's tax form. This evidence suggests there is at least *some* room for persistent individual traits, such as tax morale, to explain evasion choices across individuals. We show additional robustness checks. For instance, we show that wage under-reporting is substantial regardless of whether wages are stable or volatile, or whether the taxpayer has one or multiple employers.

We designed a survey to measure individual preferences, values, and beliefs that could explain differences in evasion choices across taxpayers. We collected *lab measures* and *survey measures*. The *lab measures* include a series of 12 laboratory games with real financial incentives. The games measure distinct individual traits that could be important for the decision to evade taxes, including honesty (the dice game), altruism (measured using the variants of the dictator game), willingness to contribute to public goods (the public goods game), aversion to tax evaders (a variant of the dictator game), among others. In turn, the 12 *survey measures* include (mostly non-incentivized) questions that are widely used in the social sciences, such as a standard question on tax morale (i.e., whether it is never, sometimes, or always justifiable to evade taxes), the perceived rate of evasion, preferences for redistribution, and trust in government. We validate these lab and survey measures by showing that, on average, the responses in our survey are consistent with those of benchmark studies. Some of the lab and survey measures we included in our study have been shown to be remarkably stable over time too (e.g. Chuang and Schechter, 2015).

From the universe of 151,565 wage earners, we invited a subset of 68,208 taxpayers to participate in the survey via email. A total of 6,078 individuals completed the survey, implying a response rate of 8.9%. The survey respondents are similar to non-respondents in all observable characteristics. Most importantly, survey respondents and non-respondents are similar in their rates of under-reporting of wages and over-reporting of deductions and withholdings. By merging the survey responses with data from administrative records, we can study whether the lab and survey measures have power to predict who are the tax evaders.

We find that the lab and survey measures have *some* power to predict who evaded taxes, but to a limited extent. Although some of the correlations between tax evasion and the lab and survey measures are statistically significant, they tend to be economically small and, in many cases, precisely estimated around zero. Some measures that we expected to be highly correlated to evasion choices, such as the survey question on tax morale, have correlation coefficients precisely estimated around zero; other measures, such as altruism, have significant correlations but not with the expected sign.

Tax morale is the most widely used variable in the literature on tax evasion, and thus we expected it to be most correlated with *actual* tax evasion. Around 24.2% of taxpayers say that it is sometimes or always justifiable to evade taxes (i.e., they have low tax morale), compared

with 75.8% who say it is never justifiable to evade taxes. The administrative data show similar rates of wage under-reporting for these two groups: 14.0% vs. 13.4%, respectively. The correlation coefficient between the survey measure of tax morale and actual tax evasion is close to zero (-0.006), statistically insignificant (p-value=0.613), and precisely estimated: according to the 95% confidence interval (CI), we can rule out a correlation below -0.032.

Some of the other lab measures that we expected to be most closely related to tax evasion have correlations that are precisely estimated around zero and are statistically insignificant. The correlation with respect to the measure of aversion to tax evaders is 0.008 (p-value=0.541), the correlation with respect to the measure of dishonesty based on the dice game is 0 (p-value=0.981), and the correlation with respect to the measure of cooperation in the public goods game is 0.003 (p-value=0.812). The only lab measures that correlate significantly with actual wage under-reporting are the three variants of the dictator game, corresponding to generosity toward a stranger, a charity, or the government, with correlation coefficients of 0.040 (p-value=0.002), 0.046 (p-value<0.001), and 0.048 (p-value<0.001), respectively. These correlations are modest in magnitude, but not negligible. The signs of these correlations, however, contradict our initial hypothesis: we find that individuals who are more altruistic are more likely to evade taxes. One potential explanation, although speculative, is that individuals who evade taxes feel the need to be generous to deal with their guilt (Gneezy et al., 2014).

The three survey measures related to wage under-reporting that have the strongest correlation and most statistical significance are the perceived evasion of firms (correlation coefficient of 0.067, p-value<0.001), left-right political ideology (-0.062, p-value<0.001), and desired tax progressiveness (0.058, p-value<0.001). These correlations indicate that individuals who are more likely to evade are those who are politically left-leaning and who think that firms evade taxes and that taxes should be more progressive. This is correlational evidence, so we must be careful with the interpretation. That being said, one speculative interpretation for this evidence is that individuals who perceive that the system is unfair find it easier to justify their own decision to evade taxes.

We provide a number of robustness checks. The correlations reported above measure evasion through an indicator variable for whether the taxpayer under-reported wages. We show that the results are similar under a host of alternative evasion measures, such as using the intensive rather than the extensive margin, and using deductions and withholdings over-reporting instead of wage under-reporting. One usual concern with survey data is that the measurement error may introduce attenuation bias. We show that the results are similar after accounting for measurement error using two standard approaches: a manual correction based on test-retest correlations, and using Instrumental Variables regressions (Gillen et al.,

2019).

Next, we quantify the power of the survey data to identify tax evaders. The area under the receiver operating characteristic (AUC, for short) measures the probability that, if you were to randomly select an individual who evaded and another who did not, the model would correctly guess which one is the evader. When considered jointly, the 24 lab and survey measures perform somewhat better than chance in predicting who under-reported wages (AUC=0.602). The model based on lab and survey measures has a bit more predictive power than a model based on the individual's evasion choices in the previous year (AUC=0.575). On the other hand, the lab and survey measures do somewhat worse than a model based on economic incentives such as marginal tax rates (AUC=0.615) and substantially worse than a model based on the individual's context as proxied by the evasion of coworkers (AUC=0.872). Indeed, this finding is consistent with a recent but growing literature documenting that contextual factors are important drivers of tax evasion.[4] Thus, although preferences and beliefs may be somewhat helpful to identify tax evaders, the taxpayer's context seems to play a much bigger role.

One potential interpretation of our evidence is that individual traits, such as dishonesty or altruism, vary little across individuals and thus are not useful to explain heterogeneity in tax evasion. Under this interpretation, factors such as tax morale can still be important to explain aggregate evasion. Indeed, even if lab measures are not good at predicting who evades more, they seem to do a decent job at predicting aggregate evasion patterns. For example, the literature on the dice game suggests that a minority of individuals act dishonestly, and even those who lie prefer not to lie all the way (Abeler et al., 2019). We find the same patterns when measuring under-reporting of wages and over-reporting of deductions and withholdings in Uruguay: only a minority of individuals cheat on these items, and even when they do, they do not take full advantage of the opportunity to cheat.

This study contributes to literature on the role of tax morale for tax compliance. The hypothesis is that differences in evasion across individuals are driven by factors such as preferences and social norms (for a recent literature review, see Luttmer and Singhal, 2014). However, there is still no consensus as to whether tax morale is a big or small factor for tax compliance. For instance, one of the first approaches to disentangle the role of tax compliance comprised cross-country regressions. In this type of analysis, it is difficult to distinguish the effect of tax morale from that of institutions. In the words of Kleven (2014), "it is not clear whether the available measures of social and cultural motives have an independent causal impact on economic outcomes, or if they are simply a byproduct of those outcomes or of

[4] For example, evidence shows that individuals learn tricks on how to evade from their peers and accountants (Chetty et al., 2013; Wilson, 2020; Battaglini et al., 2020; Boning et al., 2020; Drago et al., 2020).

deeper institutions and policies driving the outcomes."[5]

To address the concerns with causal identification, a recent literature resorts to field experiments, but that approach comes with limitations of its own. In a nutshell, these studies expose taxpayers to certain messages of "moral suasion" and then measure the effects on subsequent tax compliance (Slemrod, 2019). The evidence suggests that, while messages related to tax enforcement (e.g., tax audits) are effective at reducing tax evasion, messages of moral appeal are largely ineffective (see e.g., Slemrod et al., 2001; Blumenthal et al., 2001; Fellner et al., 2013; Castro and Scartascini, 2015).[6] The interpretation of this evidence, however, is unclear. As discussed in Luttmer and Singhal (2014), a message in a letter may be just too weak of an intervention to expect a significant impact on behavior. In other words, if an individual's tax morale is the product of decades of upbringing, culture, and education, then reading a sentence in a letter about how it is "important to do your part" is unlikely to make a significant imprint on the individual's tax morale and, therefore, on the individual's evasion choices.

We contribute to this literature by taking an entirely new approach to study tax morale. Our approach relies on the combination of individual-level data on real-world tax evasion and data on taxpayers' beliefs and values. We are the first to conduct this type of linkage for the study of tax morale. Indeed, even in economics research more broadly, there are few studies that managed to link survey data to government records at the individual level (Karadja et al., 2017; Kreiner et al., 2019; Epper et al., 2020; Hvidberg et al., 2020). Our evidence suggests that, contrary to the tax morale hypothesis, differences in evasion behavior across individuals are largely unrelated to individual preferences and beliefs. Rather, we find that economic incentives and context do a much better job at identifying tax evaders. While our evidence does not rule out that tax morale can be an important factor to explain aggregate tax evasion, it does cast doubt on the ability of tax morale to explain differences in evasion choices across individuals.

Our study is also related to literature on the relationship between behavior in the laboratory and behavior in the real world as measured with administrative data. This literature is motivated by a larger debate regarding the generalizability of laboratory experiments (Levitt and List, 2007; Camerer, 2011). Some studies show that social preferences measured in the lab have some power to predict outcomes in the field. To provide some examples related to

[5] Some attempts at causal identification include DeBacker et al. (2015), who compare the evasion of U.S. firms with ownership from more and less corrupt countries, and Halla (2012), who uses an instrumental-variable approach based on immigration data.

[6] There are some exceptions, however. There is some evidence that messages about the provision of government services may reduce tax evasion (Bott et al., 2020; Bergolo et al., 2021). There is also evidence that fairness concerns may play a role in other forms of tax compliance such as late payments (Hallsworth et al., 2017) and tax appeals (Nathan et al., 2020).

dishonesty, Hanna and Wang (2017) use data from 165 government nurses in India to show that behavior in the dice game is correlated to their real-life absenteeism. Dai et al. (2018) use data from 279 passengers of a public transport service and find a correlation between a laboratory measure of dishonesty and whether the subjects are able to display a validated ticket. And Cohn and Maréchal (2018) use data from 161 students to show that a measure of dishonesty from a coin-tossing game is correlated to school misconduct.

We contribute to this literature in two ways. First, we are able to measure real-world behavior using administrative records and in a high-stakes, natural context – indeed, cheating on taxes may be the most important form of dishonesty in economics. Second, we use sample sizes that are an order of magnitude larger than in previous studies, which is of upmost importance to address publication bias (Dellavigna and Linos, 2021). We show that the correlation between the laboratory measure of dishonesty and the real-world evasion is precisely estimated at zero. Moreover, we show that other measures of social preferences have limited power to predict real-life evasion. Indeed, our result echoes the view among some social psychologists that dishonesty behavior is rarely correlated from one context to the next (Ross and Nisbett, 2011).

The rest of the paper proceeds as follows. Section 2 describes the institutional context. Section 3 introduces the individual-level measure of tax evasion. Section 4 discusses the survey design and implementation. Section 5 presents the results. The last section concludes.

2 Institutional Context and Data Sources

In this section, we present all the key features of the institutional context that are important for the analysis.

2.1 Country Context

Uruguay is a typical country in a variety of relevant metrics, such as the size of the government, the prevalence of tax evasion, and social norms around tax evasion. Uruguay is a middle-high income country that scores low in corruption and high in human development.[7] Its tax revenue of 29.2% of the country's GDP in 2018 is similar to averages for Latin American (23.1%) and for Organisation for Economic Co-operation and Development

[7] Uruguay's annual GDP per capita was U$D 17,278 in 2018 (`https://datos.bancomundial.org/indicador/NY.GDP.PCAP.CD?locations=UY`). Transparency International ranks Uruguay 21st out of 198 countries (`https://www.transparency.org/en/cpi/2019/results`), and the United Nations Development Programme ranks Uruguay 57th among 189 countries in terms of human development (`http://hdr.undp.org/en/content/2019-human-development-index-ranking`).

(OECD) (34.3%) countries.[8] While cross-country comparisons of tax evasion are difficult, the available data suggests that tax evasion in Uruguay is not out of the ordinary. For example, according Gomez-Sabaini and Jimenez (2012), evasion of VATs was around 26% in 2008 in Uruguay, which is better than most Latin American countries and comparable to some OECD countries.[9] According to survey data, attitudes about tax evasion in Uruguay are fairly typical. For example, according to the 2010–2013 wave of the World Values Survey, 77.2% of respondents from Uruguay agree that evading taxes is never justified, compared with 68.2% for the other Latin American countries and 70.9% for the United States.

The main focus in our paper is on personal income tax (*Impuesto a la Renta de las Personas Físicas Categoria II*, or IRPF-II, in Spanish), specifically when applied to labor income.[10] Like in most of the world, the schedule of tax rates is progressive: individuals making below a minimum income have a 0% tax rate, and individuals in the highest tax bracket face a marginal tax rate of 30%. In 2016, the IRPF represented almost 13% of total tax revenue, the second-largest source of tax revenue after the VAT.[11] Like most developing countries (Jensen, 2019), personal income tax is characterized by a large income tax exemption floor. For instance, in 2016, the tax exemption threshold was set at U$D 12,436.[12] As a reference, the minimum annual wage for 2016 was U$D 5,931, and the median labor income was U$D 11,848. Consequently, the burden of personal income tax relies mainly on individuals in the middle-upper part of the gross labor income distribution (e.g., about 34% of registered workers paid personal income tax in 2016).[13]

2.2 Tax Returns and Third-Party Reporting

Each year, individuals file an annual tax return via an electronic form (1102 form), which is comparable to the 1040 forms filed by U.S. individuals and on which individuals must declare all sorts of information. Most relevant for our analysis, individuals must report all sources of earnings, including wage and self-employment incomes, as well as tax withholding and tax deductions. Form 1102 then uses those inputs to automatically compute the taxpayer's

[8] `https://www.oecd.org/tax/tax-policy/global-revenue-statistics-database.htm`

[9] Gomez-Sabaini and Jimenez (2012) reports that Uruguay has the third-lowest VAT evasion rate among the nine Latin American countries included in the study. (Gomez-Sabaini and Moran, 2014) suggest that the corresponding rate for Italy would be about 22%.

[10] The Labor income component we use does not include any type of capital income which is taxed in a separate schedule at proportional tax rates called IRPF-I. In addition to personal income taxes, Uruguay's tax revenues are collected through other taxes, such as VATs and corporate tax.

[11] Own calculations based on data from the Central Bank of Uruguay and from the Internal Revenue Service.

[12] This amount and all other dollar amounts discussed in the paper use the purchasing power parity (PPP)-adjusted exchange rate from 2018.

[13] This figure was extracted from technical reports by Uruguay's Tax Authority (`https://www.dgi.gub.uy/wdgi/page?2,principal,Documentos-Informes,0,es,0,`).

total tax burden and generate the tax due or tax refund. Mainly due to the tax exemption floor, most individuals are not required to file a tax return.[14] However, even the exempted individuals can still choose to fill a tax return, for example, if they want to claim itemized deductions that are not reported by the firm or are not subject to third-party reporting (e.g., rent or mortgage expenses). As a result of tax filing exemptions, a minority of workers file a tax return each year, but these individuals represent a substantial portion of tax revenues: e.g., in 2016, about 16% of all registered labor income earners filed a tax return, representing almost 45% of the personal income tax revenues.

As in the rest of the world, firms play an important role in individual taxation through third-party reporting. Employers must submit electronic tax form 1144 to the tax agency. This form includes information on their employees for tax purposes (similar to form W-2 in the United States). It includes wage earnings, tax withheld, and tax deductions, and it must be filed once a month.[15] Additionally, employers are required to provide their employees with an annual "income tax summary" that includes information on their total gross income, tax deductions, and tax withholdings, which employees can use to prepare their tax returns. Although this income tax summary must be provided by a deadline so that employees have enough time to prepare their tax returns, the specifics on how the information is shared are up to each firm. For example, some firms may send the information automatically to employees, and other firms may require employees to log into a website to consult the tax summary electronically.

2.3 Tax Deductions and Tax Withholding

Personal income tax is calculated based on two components: a *tax part* and a *deduction part*. The difference between these components determines the personal income tax's total liability. The *tax part* is determined by applying a progressive schedule of tax rates on the individual's gross labor income, which includes all items related to wage and self-employment income received by a taxpayer during a fiscal year. The *deduction part* includes a relatively limited set of deductions allowed by the tax code, which are also subject to a progressive deduction rate schedule. There are two types of deductions: *itemized* and *non-itemized*. *Itemized* deductions include some personal deductions, such as for child care and housing

[14] Individuals required to file a tax return include workers earning self-employment income, and individuals with wage income from multiple jobs above a designated annual threshold. Employees with a single employer are not required to file a tax return, although they can (and often choose) to do so.

[15] The 1102 form and 1144 form reported by the employer must be completed using a unique online processing software program called PARS.

expenditures.[16] This study focuses on *non-itemized* deductions, which include all social security contributions (e.g., payroll taxes and mandatory health insurance contributions) and which are proportional (i.e., a fixed rate) to the worker's gross labor income. The rate is set by law; for instance, the rate for taxpayers with dependent children is 21%.[17] Individuals who opt to file an annual tax return must report the amounts of non-itemized components to be deducted to calculate their tax liability. As previously explained, this information is available on the income tax summary provided by the employer.

Employers also play an important role in automatic income tax withholdings. Employers use a pay-as-you-earn (PAYE) tax system to withhold income taxes from monthly wage earnings.[18] Amounts withheld are treated as advance payments of personal income tax due. To calculate the amounts to be withheld, the employer uses information on monthly taxable income, non-itemized deductions, and itemized deductions claimed by the employee on the 3100 form. In practice, every month employers have to determine the employee's tax liability by computing the monthly *tax part* and *deduction part* and taking the difference between those components, as explained above. The employer is required to remit the personal income tax withholdings monthly. If the amount annual withheld exceeds the annual tax assessment, the worker is entitled to receiving a refund through either a deposit to her personal bank account or in cash.

2.4 Employee-Employer Matching of Administrative Data

We combine various sources of administrative data provided by Uruguay's Tax Authority. We construct a panel of taxpayers in 2015–2016.[19] Although we have data on self-employed individuals, our analysis focuses on individuals who earn wage income because only that group allows measuring of misreporting through their employers' third-party reports. The two main datasets are the individual tax returns (1102 form) and third-party reports from employers' statements (1144 form).[20] These datasets are matched using (masked) national identification numbers for individuals and firms. These unique identification numbers allow us to merge information from other sources of administrative data too. We thus merge

[16] *Itemized* deductions must be voluntarily claimed by the taxpayer, either indirectly through the employer using a 3100 form or directly on the annual tax return (1102 form).

[17] This 21% rate is the sum of a 15% rate for payroll taxes and 6% rate for mandatory health insurance contributions.

[18] Similar systems are used in other countries, including the United States, Australia, New Zealand, and the United Kingdom (Slemrod, 2008).

[19] In the future, we may obtain data for more recent and earlier years.

[20] We merge the tax returns and employers' third-party reports to obtain the employer and employee declarations of gross wage earnings, tax type, amount of tax deductions, and amount of income tax withheld by employers. We also have information on calculations of the *tax part*, *deduction part*, and the final tax owed (or credited back) for each taxpayer, which is available from the tax agency.

individual-level data on year of birth and gender and firm-level data on number of employees, sales, and industry activity code, among others. We also identify taxpayers' coworkers using the identification numbers of the firms.

3 Measuring Tax Evasion at the Individual Level

3.1 Sample of Interest and Descriptive Statistics

Table 1 presents some descriptive statistics. The sample of interest is comprised of 151,565 taxpayers who earned their income purely through wages and who filed a tax form in 2016.[21] For this sample, we study tax evasion by comparing the self-reported tax return to the employer's third-party report.[22] Column (3) of Table 1 presents some descriptive statistics for these 151,565 taxpayers. This sample is 44.8% female, 45 years old on average, and earns average annual gross wages of U$D 36,450.[23] It is worth mentioning that 31.5% of the individuals in this sample earned wage income from multiple employers (specifically, 25% have two jobs, 5% have three jobs, and 1.5% have 4 or more jobs). Independently of whether the employee has one, two or many employers, we can always compute the measures of tax evasion – however, when they are multiple employers, we do not have sufficient data to decompose the evasion measures separately for each employer.[24]

Table 1 also illustrates how the sample of interest is constructed and how it compares to other taxpayers in the country. Column (1) corresponds to the universe of taxpayers in Uruguay as of 2016. Comparing column (3) to column (1), the demographics (age and gender) are similar, but the sample of interest (column (3)) is on average substantially richer than the universe of taxpayers (U$D 36,450 vs. U$D 19,440). Moreover, column (2) corresponds to the subsample who filed a tax return in 2016, regardless of whether they are employees or self-employed. Comparing columns (3) and (2), the income differences between our sample of interest and the universe of taxpayers stems largely from focusing on tax filers, because only individuals making at least a certain amount are required to file a tax return.

[21] This sample excludes 24,607 individuals who, in addition to wage income, reported self-employed income. We exclude them because the comparison with third-party reports has limitations for this sample.

[22] We exclude 8,556 individuals who reported wage income on the tax return but for whom the information from the third-party report is missing in our dataset. By construction, it would be impossible for us to measure misreporting for this group.

[23] The average total gross income is slightly higher than the wage income because some individuals earn capital gains income.

[24] For instance, say that a taxpayer gets equal wage income from two employers, and she is identified as an under-reporter. Our measure of tax evasion would count her as under-reporting by, e.g., 5% regardless of whether you evaded 10% for one and 0% for the other or 5% for both.

3.2 Wage Under-Reporting

Our first, and main, measure of income tax evasion is the under-reporting of wages by employees. Taxpayers have to pay income taxes for each additional dollar that they report in wages, and therefore, they have the incentive to under-report wages to reduce their tax liability. Third-party reporting, in contexts of high enforcement capacity, should deter employees from under-reporting income for tax evasion (Kleven et al., 2011, 2016). For instance, in the United States, if an employee under-reports the wage relative to the employer's third-party report, the Internal Revenue Service would automatically correct the individual's tax form, update the tax amount due, and then notify the taxpayer. Wage income subject to third-party reporting is heavily enforced in developed countries, such as the United States (Internal Revenue Service, 2016) and Denmark (Kleven et al., 2011). In developing countries, however, the effect of third-party reporting on tax compliance is constrained due to a weaker administrative capacity (Carrillo et al., 2017), even under the presence of tax withholding at the source (Brockmeyer and Hernandez, 2016).[25] In the country we study, Uruguay, the personal income tax had been introduced recently (in 2008). Thus, the tax administration was still dealing with all sorts of pressing issues with its implementation. We study a specific time period (2015–2016) during which the tax authority did not conduct a systematic and automatic cross-check between the individual tax returns and the third-party reports.[26] As a result, employees could under-report wages to evade taxes during our sample period.[27]

Figure 1.a presents the results for the discrepancies in wage reporting in 2016. Around 80.6% of wages reported by individuals coincide perfectly or almost perfectly (i.e., within 1%) with the third-party report of the employer (black bar). Around 15.5% of individuals under-report their wages (red bars). That is, the reported wage is at least 1% below the employer's third-party report. Among those who under-report, the average under-reporting is 6.1%. These discrepancies could reflect honest mistakes. However, we provide robust evidence to the contrary. Figure 1.a shows that income under-reporting is much more likely than over-reporting (blue bars): 15.5% of taxpayers under-report to some extent, but only

[25] Third-party information also has limits as a compliance mechanism, even in environments with high tax enforcement, especially if taxpayers can adjust in margins that tax authorities cannot easily verify (Slemrod et al., 2017).

[26] There was a major change in 2017, when the tax authority introduced pre-filled tax returns. On the pre-filled return, items such as labor income or personal deductions of taxpayers are pre-filled based on information from third-party reports, which may reduce this form of tax evasion by affecting the perceived probability of detection by taxpayers or by making unethical behavior more salient. Unfortunately, we do not have access to data after 2016 as of now to test this conjecture.

[27] As in other forms of tax evasion, the evasion attempt could backfire. For example, if they were selected for an audit, the discrepancy with the third-party report could be detected, in which case the employee would have to pay the amount evaded as well as the corresponding fines. The audit probabilities and penalty rates, however, tend to be quite low (Bergolo et al., 2021).

3.9% over-report their wages. The fact that taxpayers make disproportionate mistakes in the direction that best serves their financial interest suggests that under-reporting is largely intentional.

That figure also shows that the significant fraction of individuals who under-report their wages do not do it excessively: the median under-reporting is 4.8%, and the 25th and 75th percentiles are 3.1% and 7.0%, respectively. One potential interpretation for this behavior is that individuals engage in "motivated errors": i.e., subjects may act as if they are making a "mistake" that reduces their tax burden (Exley and Kessler, 2019). For instance, employees may be rounding down their true salary, accidentally reporting the wage from the previous year, or "forgetting" to include a bonus payment or sales commission.

To put these magnitudes in perspective, we estimate how wage misreporting affects the tax liability. To that purpose, we estimate the amount of tax misreported (i.e., the tax gap) as the difference between the individual tax liability based on actual tax returns versus what it would have been if, holding constant all the other items in the tax form, they had reported exactly the wages reported by the employer. Figure 1.b shows the distribution of the tax gap relative to the "true" tax liability. According to this analysis, 15.7% of individuals evade taxes (i.e., they pay less in taxes than they should).[28] Moreover, conditional on evading something, they evade an average of 17% of their tax liability, which corresponds to U$D 344.[29]

Finally, we can benchmark our results to other results from Pakistan showed in Best (2014) in 2007–2012. Similar to our study, Best (2014) compares third-party reports to the wages reported by the taxpayer in the tax form and classifies them as under-reporters if they sub-declare by 0.25% or more. Best (2014) reports that 19.3% of employees from Pakistan under-report wages and, conditional on doing so, they understate them by an average of 15.6%. In comparison, using the same 0.25% threshold, we find that 18.5% of Uruguayan employees under-report wages by 5.2%, on average. The comparison between the Uruguay and Pakistan contexts has several caveats, however, including differences in institutional contexts, subject pools, and even how data is collected and processed. To the extent that tax compliance tends to be higher in more developed countries, we arguably expect tax compliance to be higher in Uruguay than in Pakistan. Thus, it is reassuring that despite all of those differences, the extent of wage under-reporting reported for Uruguay is in the same order of magnitude reported by Best (2014) for Pakistan.

In contrast, the extent of wage under-reporting should arguably be much lower in developed countries than in developing countries. Factors such as automated tax filing systems

[28] We define taxpayers as tax evaders if their tax gap exceeds 1%.
[29] The median is 11% with an inter-quartile range of [6.9%, 18.4%].

(e.g., pre-populated tax returns) and automatic cross-checking should make it difficult for employees to under-report wages relative to the third-party reports. Consistent with that conjecture, we observe that the degree of wage under-reporting seems to be much lower in some developed countries. For example, Kleven et al. (2011) uses data from audits and reports that the fraction of wage under-reporters was 1.4% in Denmark in 2006, which is an order of magnitude lower than the levels of wage under-reporting in our data for Uruguay. The evidence suggests that wage under-reporting is rare in the United States too (Johns and Slemrod, 2010).[30]

3.3 Deduction Over-Reporting and Withholding Over-Reporting

Albeit suggestive, this evidence is not proof that individuals under-report their wage earnings to evade taxes. This under-reporting instead could be due to asymmetrical mistakes (i.e., employees may be more likely to make downward errors than upwards errors). We test this hypotheses by exploiting the institutional context: although employees have incentives to under-report wages, they also have incentives to over-report tax deductions and tax withheld, which are both third-party reported. In the case of deductions, our analysis focuses on the subset of non-itemized deductions (i.e., items that are proportional to the worker's gross labor income). For the sake of brevity, in the rest of the paper we refer to non-itemized deductions as just tax deductions.

Figure 1.c shows the results for deduction discrepancies. For the sake of simplicity, we start with the subjects who report wages within 1% of the third-party report. The results on deduction discrepancies from Figure 1.c are consistent with the results on wage discrepancies from Figure 1.a: most individuals correctly report their deductions, but those who misreport do it disproportionately in the direction that reduces their tax bill. More specifically, whereas 15.8% of individuals over-report their tax deductions, only 2.2% under-report them. Figure 1.d is similar to Figure 1.c but focuses on misreporting of tax withholdings instead of tax deductions. Like with tax deductions, but contrary to wages, individuals have an incentive to over-report them. Figure 1.d shows that a significant fraction (22.7%) of individuals over-report their tax withholdings, but a much smaller share (2.9%) under-report tax withholdings. In sum, the evidence shows that individuals make a disproportionate number of "mistakes" in the direction that reduces their tax liability.

[30] Johns and Slemrod (2010) use data from a 2001 sample of audited taxpayers from the Internal Revenue Service national research program and report that the wages subject to third-party reporting and tax withholding were under-reported by an average of just 1%.

3.4 Evasion Choices Across Years

One interesting feature of our data is that we observe the same taxpayer making multiple decisions: each taxpayers must choose whether to lie in a specific item of the tax return (e.g., under-report wages), and then faces the same choice a year later. Whether the decision to evade is persistent can provide suggestive evidence on the underlying roots of tax evasion. Indeed, we find that individuals who cheat one year are significantly more likely to cheat the following year. Figure 2 shows the same evasion decision (wage misreporting) for the same individual but over different years. This sample is based on the subset of individuals who filed tax returns and were pure wage earners in both 2015 and 2016. Figure 2.a corresponds to the subsample (79.6%) of individuals who reported wages accurately in 2015 according to third-party reports, and Figure 2.b corresponds to the subsample (10.6%) who under-reported their wage in 2015. The data show a significant persistence in these choices. For example, among accurate reporters in 2015, 15.2% under-reported wages in 2016; among under-reporters, however, a substantially higher share (26.5%) under-reported wages the next year. The difference between these two is large, 11.3 percentage points (pp), and highly statistically significant (p-value<0.001). The persistence across time indicates that evasion choices are not purely random. Moreover, this evidence suggests that there is *some* scope for persistent individual traits, such as tax morale, to explain evasion choices across individuals.

3.5 Other Robustness Checks

For some employees, wages can increase or decrease from one year to the next. For other employees, wages can remain the same. It is possible that these wage changes play a role in the wage misreporting. For example, perhaps employees are under-reporting wages because they forget, conveniently or accidentally, to update their reported wage to reflect recent raises. Or perhaps employees who have stable incomes do not misreport because there is less room for "mistakes." Figure 3 presents a simple analysis to explore whether wage changes play a role in wage misreporting. Each bar represents the average rate of wage under-reporting in 2016 for a different group of employees. The employees are divided in groups (listed in the x-axis) according to the change in their third-party reported wages from 2015 to 2016.[31] For example, the leftmost group corresponds to employees whose wages declined by 20% or more, while the rightmost group corresponds to employees whose wages increased by 20% or more. While there are some differences across groups, the rates of wage under-reporting are always in the same order of magnitude regardless of whether wages went up, down or stayed

[31] As in the analysis of persistence shown above, this sample is based on the subset of individuals who filed tax returns and were pure wage earners in both 2015 and 2016.

the same year-over-year.

Just like wage volatility could create more room for misreporting, it is possible that having multiple employers could be conducive to under-reporting. Perhaps taxpayers are reporting truthfully the wages from their main employer and failing to report the income from secondary employers. Or perhaps taxpayers have a harder time keeping track of their wages when they have multiple employers. To explore this mechanism, we split the sample in taxpayers who have a single employer vs. taxpayers with more than one employer. We find that the rates of under-reporting are in the same order of magnitude in the two groups (results reported in Appendix A.1). This evidence indicates that having multiple employers may play a role, but a minor one.

It is important to note that our measure of tax evasion is not the only form of tax evasion related to wages, and thus constitutes a lower bound to tax evasion. Most importantly, employers and employees may collude to have part of the compensation off the books and thus reduce the burden of payroll taxes or social security contributions linked to their employee's wages (and presumably split the savings between the two). There is some recent evidence that this type of collusion may take place, specially in small firms (Bjørneby et al., 2021; Biro et al., 2020). Our measure of tax evasion takes the third-party reported wage for the employer as given, and then studies the choice of the employee to report that same amount or something different. As a result, our wage under-reporting measures the amount of tax evasion above and beyond any additional evasion that may (or may not) take play through collusion.

Since we are measuring the gap between the wages reported by the employer and the employee, one potential concern is that employees may be reporting truthfully while employers may be the ones misreporting. For instance, employees may be reporting how much they truly got paid from the employers, while employers may be over-reporting that amount to the tax agency. This alternative interpretation is not even brought up in related studies (Kleven et al., 2011; Best, 2014; Bergolo et al., 2019), probably because firms do not have a clear incentive to over-report wages to the tax agency.[32] Moreover, it is unclear why firms would do this type of over-reporting for some of their employees but not for their other employees. In any case, we provide direct evidence against this confounding factor, by focusing on the deductions misreporting instead of the wage misreporting. We interpret the discrepancies between the deductions reported by the employee vs. the employer as evidence that employees are conveniently over-reporting deductions to evade taxes. The alternative interpretation

[32] While this over-reporting could reduce the tax burden of the firm by inflating costs, on the other hand it would create additional, and likely greater, tax burden through the higher payroll taxes and social security contributions linked to the employees' wages. If the employers wanted to reduce their corporate income tax, they would be better off by inflating other expenditure items besides wages.

would be that employers are under-reporting their employees' deductions. However, firms would have no incentive to under-report the employee's deductions: whatever tax deductions their employees are getting does not affect the amount that the employer has to pay in taxes.

4 Survey Data

4.1 Survey Design: Overview

The survey is designed to be implemented with a sample of taxpayers who recently filed a tax return. The English translation of the full survey instrument is included in Appendix C, and the original Spanish version is in Appendix E. This survey was pre-registered in the Registry for Randomized Controlled Trials operated by the American Economic Association (RCT ID #0004108). The survey starts by collecting some background and demographic information about the respondent and then elicits the lab and survey measures, described in detail in the following sections.

4.2 Lab Measures

The respondent completes a series of laboratory games implemented as a series of incentivized survey questions. Right before starting, respondents see a screen explaining how the incentives work and emphasizing the importance of answering carefully and honestly because the games offer a real financial incentive: upon completion of the study, 50 respondents will be randomly selected to have one of their choices "executed". In other words, for each participant, one of their incentivized decisions will be randomly chosen, and the payouts will be calculated based on that decision. For example, in the dictator game, if the first player chooses to split $U1000 equally between herself and the second player, the researchers would pay $U500 to each of them.[33] For those who are not among the 50 chosen respondents, their decisions remain hypothetical. The method of "executing" a random sample of choices is a common feature in laboratory experiments. Moreover, there is direct evidence that the probability with which each choice is "executed" does not matter as long as it is positive (Carson and Groves, 2007; Charness et al., 2016). To make the real financial consequences more salient, the following reminder message is displayed at the bottom of the screen for each of the incentivized questions: "There is a chance that this decision will be executed and therefore your choice could have real consequences."

[33] Most games involve two players, but some games have more than two players. In all games, one or two players make decisions that affect the final allocation to two or more players. Games for which more than one player has choices are played sequentially. In all cases, the players and their choices are anonymous.

We include the following adaptations of well-known laboratory games that are designed to measure specific traits (e.g., altruism, honesty) that could affect the decision to evade taxes.

- **Evasion aversion:** We designed this game specifically to measure attitudes towards tax evasion in an incentive-compatible way. In this game, the subject (player A) decides how to assign money to two random taxpayers (player B and player C). Player A learns that player B thinks it is never acceptable to evade taxes and that player C thinks it is sometimes acceptable to evade taxes. Player A then decides how much of $U1,000 to give to player B (with the remainder going to player C), anywhere from $U0 to $U1,000 in $U250 increments. Regardless of the decision, player A receives $U1,000. The share of the endowment given to player B, who thinks evading taxes is not justified, constitutes our incentivized measure of aversion to tax evaders. The hypothesis is that individuals who are more averse to tax evaders will be less likely to evade taxes.

- **Public good:** In the literature on tax compliance, the decision to evade taxes is typically modeled as the decision to become a free-rider in the context of a public good provision (Cowell and Gordon, 1988). Thus, we include a simple variant of the public good game. Player A is paired with four random taxpayers. Each of the five players must decide how much of their $U1,000 endowment to contribute to the public good, $U0 to $U1,000 in increments of $U250, knowing that the total contribution will be doubled and then divided equally among the five participants. The share of the endowment contributed to the common pool measures the willingness to cooperate in the provision of public good. The hypothesis is that individuals who contribute more to the public good will be less likely to evade taxes.

- **Dishonesty:** To check whether dishonest people evade taxes, we included a classic lab measure of honesty: the dice game (Fischbacher and Föllmi-Heusi, 2013). The respondent is asked to report the outcome of a (private) die roll and then receives a reward that is proportional to the number reported, thus incentivizing to over-report the number rolled. The hypothesis is that individuals who are more likely to lie in the dice game are also more likely to evade taxes. Moreover, to serve as a benchmark, we include a game where individuals can earn more by lying, but they have to lie about information that is verifiable: we ask subjects whether they were born in an even or odd year and explain that we will pay $U500 if they answered an odd year or $U2,500 if they answered an even year. Since we observe the year of birth in the administrative records, we can measure directly whether the subjects are lying or not. The hypothesis

is that individuals would be much less likely to lie about information that is ex-post verifiable such as the year of birth.

- **Giving to strangers:** Another reason why individuals may choose not to evade taxes is that they do not want to harm the individuals who benefit from the services that are financed through those tax revenues. Thus, we include three variants of the dictator game to measure the respondent's generosity towards groups that can benefit from tax revenues. In this first variant, the respondent chooses how much of a $U1,000 endowment to share with a stranger, on a scale from $0 to $U1,000 in increments of $U250. The hypothesis is that individuals who are more generous will be less likely to evade taxes.

- **Giving to charity:** One could argue that the tax revenues do not benefit a random stranger but are disproportionally likely to benefit the neediest in the population. In this game, respondents can donate part of their endowment to a well-known nonprofit organization that provides charitable education services. Players choose how much of their $U1,000 endowment to donate on a scale from $U0 to $U1,000 in increments of $U250. The hypothesis is that individuals who are more generous will be less likely to evade taxes.

- **Giving to government:** Some individuals may want to help people in need but do not want to pay taxes because they believe that those tax dollars are spent inefficiently (Alm et al., 1992, 2012). For this reason, we include yet another variant of the dictator game in which the respondent can donate part of their endowments to government or to a nonprofit organization. To make it comparable to the donation to the charity, the recipient of this donation is a government agency that provides education services comparable to those provided by the charity. Respondents choose how much of their $U1,000 endowment to donate on a scale from $U0 to $U1,000 in increments of $U250. The hypothesis is that individuals who are more generous towards the government will be less likely to evade taxes.

- **Trust:** An honest taxpayer may mistrust others and suspect that they evade, which results in the typical free-rider problem (Alm et al., 2012; Fehr, 2009). To capture this phenomenon, we include the standard trust game (Berg et al., 1995). The subject (player A) decides whether to invest all or none of a $U1,000 endowment in another random player (player B). If player A does not invest, then both players get $U1,000 each. If player A invests, then player B receives $U4,000 and must choose how to split the earnings (i.e., both receive $U2,000 or player B keeps the whole $U4,000). Choosing

to invest would indicate that the subject is willing to trust others. The hypothesis is that individuals who are more trusting will be less likely to evade.

- **Ultimatum:** Some studies argue that the decision to evade taxes may be related to social preferences, such as fairness and equality concerns (Alm et al., 1995; Andreoni et al., 1998). We thus include a series of games related to social preferences. First, we include a simple version of the ultimatum game. The respondent (player A) is paired with player 2 (a random taxpayer), who proposes how to split a $U1,000 endowment: $U800 for player B and $U200 for player A. Player A may accept or reject the offer (in which case both players get nothing). The choice to reject the offer measures whether the subject cares about fairness. The hypothesis is ambiguous: depending on whether the individual thinks that tax evasion increases or decreases fairness, the decision in the ultimatum game may be positively or negatively correlated to tax evasion. An individual who believes to have been treated unfairly by society may see tax evasion as an opportunity to make things fair. On the other hand, an individual could see tax evasion as intrinsically unfair to fellow citizens. Indeed, the hypothesis is ambiguous not only for this ultimatum game but for all the following games regarding social concerns.

- **Inequality aversion:** In this game, the respondent (player A) must choose between two possible allocations for two random taxpayers (players B and C). Player A can choose $U250 to player B and $U250 to player C, or an unequal split of $U250 to player B and $U750 to player C. Regardless of the decision, player A receives $U1,000. Choosing an equal split indicates that the player cares more about equality than about efficiency. The hypothesis is, again, ambiguous: individuals who care more about equality may be more or less likely to evade.

- **Meritocratic preferences:** According to a meritocratic fairness view, more productive workers should receive higher income than less productive ones (Almås et al., 2020). In this game, the subject (player A) decides how to split an endowment of $U1,000, in increments of $U250, between two random taxpayers (players B and C). Player A learns that player B performed a simple task on the computer for 15 minutes while player C did nothing. Giving a higher share of the endowment to player B (the one who worked) would reveal that player A is more willing to tolerate inequality that comes from "effort" rather than from "luck," which is consistent with a meritocratic view (Almås et al., 2020). The hypothesis is again ambiguous: individuals with a meritocratic view may be more or less likely to evade.

- **Impatience:** Because the potential costs of evading taxes (e.g., fines, lower provision

of public goods) happen in the future, it is possible that less patient individuals are tempted to evade more. We measure impatience using the "staircase" procedure for intertemporal choice proposed in Falk et al. (2018): the subject makes up to five sequential choices that measure his or her willingness to accept delaying a payment for a year (Appendix A.2 presents the full decision tree). The hypothesis is that more impatient individuals will be more likely to evade.

- **Risk aversion:** The standard model of tax evasion views it as a risky investment (Allingham and Sandmo, 1972). In this view, evading is akin to a risky lottery that has a high probability of paying out the evaded amount but also a certain probability of being costly (i.e., paying fines) if caught. Thus, the decision to evade may be mediated by the individual's degree of risk aversion. We measure risk aversion using the staircase procedure proposed in Falk et al. (2018): the subject must makes up to five sequential choices that allow us to identify the certainty equivalent for a risky lottery (Appendix A.2 presents the full decision tree). The hypothesis is that more risk-averse individuals will be less likely to evade.

4.3 Survey-Measures

We include a series of questions used in social science to measure values and beliefs (e.g., tax morale, preferences for redistribution) that may predict the decision to evade taxes. Each measure is summarized briefly below:

- **Tax morale:** We ask, "How justifiable do you think it is to evade taxes?". Responses range from "never," "sometimes," and "always". This type of survey question is the most widely used in the literature on tax morale (Torgler, 2005; Cummings et al., 2006; Frey and Torgler, 2007; Halla, 2012). The hypothesis is that individuals who think it is sometimes or always justifiable to evade taxes will be more likely to do so.

- **Workers' evasion:** We use a simple measure of descriptive social norms. After a brief explanation on how employees may under-report wages, we ask individuals to guess the percentage of employees who under-report their wages using bins from "0-10%" to "90-100%". To encourage honest guesses, we include this and the following question as part of the incentivized games. Subjects are told that we will compare their guesses to the results from a recent academic study, and if they choose the correct option, they could win $U1,000. The hypothesis is that individuals who perceive high rates of wage under-reporting will be more likely to under-report themselves, presumably due to weaker perceived social norms.

- **Firms' evasion:** In addition to perceptions about the evasion rates among employees, we also elicit the perceived evasion rate among firms. Since the VAT is the largest source of taxation for firms, we ask respondents to guess the average VAT that companies under-report. As with the previous question, we provide a potential $U1,000 reward for an accurate guess. The hypothesis is that individuals who perceive high evasion rates among firms will be more likely to evade taxes themselves, due to social norms or perhaps fairness concerns.[34]

- **Trust in others:** In addition to measuring interpersonal trust with a laboratory game, we also include the standard attitudinal survey question in a 3-point-scale used in the American General Social Survey and the World Values Survey: "Generally speaking, would you say that most people can be trusted, or that one can never be careful enough when dealing with others?" The hypothesis is that individuals who are more trusting will be less likely to evade taxes.

- **Trust in government:** Compared with interpersonal trust, trust in the government may be more important to foster tax compliance (Feld and Frey, 2002). We thus include the following question: "Would you say that the government can generally be trusted to act correctly?" The responses on a 5-point scale range from "never" to "always." The hypothesis is that individuals who trust the government more will be less likely to evade taxes.

- **Government efficiency:** Individuals may trust government but think that tax revenue is wasted due to inefficiency. We thus include the question: "Do you think that the government is efficient in the way it manages public resources?" The responses on a 4-point scale range from "very inefficient" to "very efficient." The hypothesis is that individuals who think the government is efficient will be less likely to evade taxes.

- **Preferences for redistribution:** Taxation plays a key role in the provision of public goods and redistribution of income. It is therefore possible that individuals want to pay their taxes only if they agree with the goal of redistributing income (Castañeda et al., 2020). We elicit the respondent's agreement with the following statement on a 4-point scale from "totally agree" to "strongly disagree": "Governments should take steps to reduce the income gap between rich and poor." The hypothesis is that individuals who support income redistribution should be less likely to evade taxes.

[34] Moreover, after eliciting these prior perceptions about workers and firms, we embed an information-provision experiment to create exogenous variation in the posterior beliefs. Unfortunately, due to unforeseen challenges we have not been able to access the administrative data for the post-survey period and thus we cannot estimate the effects of the experiment yet.

- **Left-Right Spectrum:** It is possible that individuals do not want to evade taxes for ideological reasons. For example, individuals who believe that taxation is ethically wrong may not want to pay their taxes (Doerrenberg and Peichl, 2018). We use the standard measure of a public opinion research question based on self-reported positions on an 11-point scale with a middle point of 5: "In politics, we usually speak of *left* and *right*. On a scale where 0 is the left and 10 is the right, where would you be located?" The hypothesis is that individuals towards the right of the political spectrum may be more likely to evade taxes.

- **Perceived progressiveness:** Perhaps individuals do not want to pay taxes because they do not agree with how the tax burden is distributed in the population. We adapt a question used in Kuziemko et al. (2015) that elicits the average tax rate for three steps of the economic ladder in the country: "In 2017, what percentage of their gross personal income do you think that the following social groups actually paid in personal taxes, on average?" We elicit the tax rate in the lower class (the bottom 20% of the income distribution), the middle class (the middle 60% of the income distribution), and the higher class (the top 20% of income distribution). For each group, the responses are a continuous variable ranging from 0% to 100%. For reference, we inform respondents that the average tax rate for the whole population is 21%. We define perceived progressiveness as the difference between the perceived tax rate paid by the higher class and the lower class. Based on the model by Doerrenberg and Peichl (2013), the hypothesis is that individuals who perceive a more progressive tax schedule will be less likely to evade.

- **Desired progressiveness:** Perhaps what matters the most is not whether individuals think that the tax schedule is currently progressive but whether they think it should be more progressive or less progressive. We measure the designed progressiveness by eliciting the respondent's agreement with the following statement on a 4-point scale (from "totally agree" to "strongly disagree"): "Tax rates should be more progressive (that is, higher for the rich and lower for the poor)".[35] According to the model by Doerrenberg and Peichl (2013), the hypothesis is that individuals who desire a more progressive tax schedule will be less likely to evade.

- **Perceived inequality:** We include a lab game to assess attitudes towards inequality. To complement that measure, we elicit the tolerance for inequality using a question that is widely used in the literature on preferences for redistribution: "What do you think

[35] This is a simplified version of a question used in Kuziemko et al. (2015) to measure the ideal tax rate.

about income differences between the rich and the poor in Uruguay?" The responses use a 3-point scale ("too low," "about right", or "too high"). The hypothesis is again ambiguous: individuals who care more about equality may be more or less likely to evade.

- **Role of Luck:** To complement the lab game that measures inequity concerns, we include a question adapted from Kuziemko et al. (2015) asking whether luck or effort is more important to explain why some individuals are poor and some are rich. The response takes the value 2 if luck is important for being rich and poor, 1 if luck is important for one of the two, and 0 if it is important for neither. The hypothesis is ambiguous: individuals who think luck is more important may be more or less likely to evade taxes.

4.4 Implementation Details and Descriptive Statistics

We sent invitations by email between April 2019 and June 2019, and all responses were collected during that same time window. Invitations were sent to the email addresses that taxpayers reported to the tax agency.[36] Appendix B includes an English translation of the invitation email, and Appendix D includes the original Spanish version. This invitation describes a short survey for academic purposes conducted by researchers from universities in Uruguay, Argentina, and the United States. The invitation mentions that the survey relates to economic opinions and attitudes but provides no specifics on the hypotheses being tested. The invitation mentions a small monetary incentive to participate in the 20-minute survey: 20 raffle prizes of U$D150 each (plus additional potential rewards from the incentivized games).[37] The invitation also emphasizes that participation in the survey is voluntary, that the responses to the questionnaire are confidential and would be used only for academic purposes. To comply with this promise, after the survey responses were linked to the administrative records from the tax agency, the individual identifiers were removed. As a result, the survey responses remained anonymous to both the tax agency and the researchers.

Table 1 provides descriptive statistics about the survey respondents and non-respondents. Column (3) corresponds to the relevant sample used for the analysis in Section 3 (i.e., the wage earners who filed a tax return).[38] Columns (4) and (5) further decompose that sample

[36] Taxpayers must file their tax returns in electronic format and submit the file by e-mail or in person at an official tax office. During the analysis period, including an e-mail address in the tax return was optional. Taxpayers were required to include a postal address and phone number only. Therefore, we collected email addresses for taxpayers who either included their email address in the tax return or delivered their tax returns by email (i.e., as an attachment).

[37] The median respondent took 21.88 minutes to complete the survey.

[38] We sent invitations to individuals outside of this sample, including 21,980 taxpayers with self-employment

into individuals who were invited and not invited to the survey. Of the 151,565 taxpayers who are pure wage earners and filed a tax return, we invited 68,208 to participate in the survey study. Taxpayers who were not invited did not have a valid email address registered with the tax agency.[39] Of the 68,208 invited, 6,078 completed the whole survey, or a response rate of 8.9%.[40]

Table 1 compares the characteristics of individuals who responded to the survey (column (7)) with those who were invited to the survey but did not respond (column (6)). Due to the large sample sizes, the differences in average characteristics between the two samples are almost always statistically significant. However, the differences between survey non-respondents (column (6)) and respondents (column (7)) are always small in magnitude (e.g., 52% vs. 58% in the share of women, 44 years vs. 43 years in average age, and U$D42,000 vs. U$D41,000 in annual income.[41] This evidence suggests that responding to the survey is largely orthogonal to the observable characteristics.

One potential concern is that individuals who evade taxes are the type of individuals who would not respond to surveys. To assess this possibility, the last three rows of Table 1 report the three measures of tax evasion: under-reporting of wages and over-reporting of deductions and withholdings. The comparison of columns (6) and (7) indicates that, contrary to the potential concern, whether an individual responds to our 2019 survey is largely unrelated to their tax evasion in 2016. For example, the fraction of taxpayers who under-reported wages in 2016 is 13.5% among survey respondents (column (7)) versus 13.8% among non-survey respondents (column (6)). This finding is robust if, instead of comparing the fraction of wage under-reporters, we use the fraction of deduction over-reporters or the fraction of withholding over-reporters.[42]

Last, we provide some direct evidence that the respondents paid close attention and

income, 1,828 of whom completed the survey. However, we do not present the data here because so far, we do not have a measure of tax evasion for the self-employed and thus they play no role in the current version of the study.

[39] We excluded invalid email addresses, such as those without an "@", and e-mail addresses that appeared more than once, which likely belonged to a preparer other than the taxpayer, such as a family member or accountant. In addition to the first invitation email, subjects who did not complete the survey were sent a reminder email, typically a week after the original email.

[40] A survey was considered complete if the respondent completed 100% of the questions. Of the 7,858 individuals who clicked on the link to the survey and advanced to the first screen, 6,078 finished the whole survey and 1,780 finished part of it. In the current version of the study, we exclude partial responses from the analysis.

[41] Differences between the universe of taxpayers (column (3) of Table 1) and the survey respondents (column (7)) arise entirely from differences between individuals invited and not invited to the survey (columns (4) vs. (5)).

[42] The small differences in the degrees of misreporting between the survey sample (column (7) of Table 1) versus the universe of wage earners (column (3)) are explained almost entirely by the differences between individuals who were invited or not invited to the survey (columns (4) vs. (5)).

understood most of the survey questions. We included one question at the end of the survey to assess the clarity of the survey, and 98% of the respondents reported that they understood every or almost every question of the survey. We used two methods to check if respondents paid close attention to the instructions. First, at the end of the questionnaire, we added a long question as an attention check, asking the respondent to select one specific option among a large number of options, and 94% responded correctly. This rate is high and consistent in magnitude with other survey studies using this same question.[43] Second, we purposely included a question asking the subject's gender and then checked their responses against the administrative data. Almost all (95.9%) survey respondents reported the same gender as in the administrative records. In sum, these results indicate that subjects paid close attention to the questions and did not complete the survey quickly to be eligible for the raffle prizes.

4.5 Variation in Lab Measures

Figure 4 presents the raw distribution of responses for each of the 12 lab measures used in this study. Each question shows variation in the responses across individuals. For example, Figure 4.d shows that in the dictator game, around 24.1% of respondents share 0% of the endowment with their partners, 21.0% of respondents share 25%, 50.9% share exactly half of the endowment, and 4.0% share more than half. The question with the least variation is the inequality aversion game (Figure 4.i), in which most subjects (85.3%) choose an even split.

One game that deserves special attention is the dice game. Under the null hypothesis that everyone is honest, we would expect one sixth of the respondents to fall into each option 1 through 6. Figure 4.c, however, shows that this is not the case: while the highest number (6) is reported almost exactly one sixth of the time, low numbers (1 and 2) are reported less frequently than one sixth of the time and medium to high numbers (3 to 5) are reported more frequently than one sixth of the time. Indeed, this distribution of responses mimics closely what has been reported in other studies based on different populations. The intuition is that while we can reject that everyone responds honestly, most people seem to be honest. Moreover, when people are dishonest, they seem to avoid taking full advantage of the lie (i.e., lying about rolling a 6). We summarize these results by associating the "excess mass" in numbers 3 to 5 to the probability of lying. For example, 22.6% of respondents pick answer 4, but we only expect 16.7% to actually get that number by chance, so the probability of having lied conditional on reporting number 4 is 26.3% ($= \frac{22.6-16.66}{22.6}$). According to this rough estimate, about 11% of the subjects lied in the dice game. Recall that we include an additional

[43] For example, 96.4% of respondents passed this attention check in Bottan and Perez-Truglia (2020). This high accuracy is even more remarkable given that it is included at the end of the survey when respondents may be subject to survey fatigue.

game to use as a benchmark, in which we give individuals incentives to lie about something that is ex-post verifiable (whether their year of birth is an even or odd number). Although individuals who were born on an even year had a strong financial incentive to misreport that they were born on an odd year, only 3% of respondents do so (for more details, see Appendix A.3). This result is consistent with findings in the literature that individuals are much less likely to lie about things that can be observed and verified by others (Crede and von Bieberstein, 2020).

We also find that the variation between different lab measures are mostly orthogonal to each other. For example, the pairwise correlations between lab measures range from -0.140 to 0.260, with an average correlation of 0.016 (see Appendix A.3). In other words, rather than measuring the same trait repeatedly, these games seem to measure different features of individual preferences.

We also provide a validation test for these lab measures. We benchmark the average choices in these games against the average responses reported in (arguably) similar laboratory experiments. Figure 6.a presents the results for all 12 lab measures. For ease of interpretation, the variables are constructed to take values from 0 to 1.[44] For example, the variable for giving to strangers is equal to the fraction of the endowment that the dictator shares with a stranger, with 0 corresponding to nothing and 1 to everything.

The results from Figure 6.a indicate that the lab measures line up reasonably well with the corresponding measures from other studies. For example, the average subject in our dictator game shares 34.6% of the budget with a partner. In comparison, Engel (2011) reports that, on average, dictators in their study shared 28.3% of the budget. Regarding the dice game, 11% of respondents lie in our sample, compared to 21.8% in Gächter and Schulz (2016). More generally, the average choices in our experiment are significantly correlated to those in the benchmark studies (correlation coefficient of 0.763, p-value=0.004). Note that we should not expect the average behavior to be identical within each pair of studies. For example, there are differences in population type (e.g., random sample of taxpayers from Uruguay versus undergraduate students from the United States), the stakes involved, and the language and framing of the game. Despite these differences, it is reassuring that the average behavior in our sample is largely consistent with the average behavior in the literature.

4.6 Variation in Survey Measures

Figure 5 presents the raw distribution of responses to the 12 survey measures, again showing ample variation across individuals. For example, in response to the question about their

[44] The only exception is the variable *impatience*, which can take values from 0 to 1.2. See the notes to Figure 6 for a full list of variable definitions.

tax morale (Figure 5.a), 75.8% of subjects report that it is never justifiable to evade taxes, compared with 22.7% who report that it is sometimes justifiable and 1.5% who respond that it is always justifiable. We again observe that the variations between these different lab measures are mostly orthogonal to each other, with pairwise correlations ranging from -0.47 to 0.61 and an average correlation of 0.04 (see Appendix A.3 for more details). In other words, these survey questions seem to measure different features of individual preferences and beliefs.

We further validate the survey measures by benchmarking them to other sources of survey data from Uruguay. Figure 6.b presents this comparison for 10 survey measures in our study for which we could find a benchmark. As in Figure 6.a, Figure 6.b defines the survey measures to take the values from 0 to 1, for ease of exposition. The results from Figure 6.b show that, again, the average responses to the survey questions in our sample line up reasonably well to the corresponding responses in other Uruguayan surveys. For example, in our question of tax morale, 75.8% of subjects report that it is never justifiable to evade taxes. In the 2011 World Values Survey for Uruguay, responses to the same question reveal a similar order of magnitude: 76.5% of respondents declared that it is never justifiable to evade taxes. The correlation between the average survey responses in our experiment and the benchmark studies is quite significant (0.76, with a p-value=0.011). The distribution of responses is unlikely to be identical due to differences in how the surveys were implemented and how respondents were recruited. For example, the benchmark surveys tried to recruit a representative sample of the whole country, whereas our survey was directed towards individuals who file tax returns and thus tend to belong to the upper echelons of the income distribution. However, despite these differences, it is reassuring that the responses are largely consistent across surveys.

5 Predicting Evasion Choices

5.1 Relationship Between Tax Evasion and Selected Measures

We leverage the link between the survey and administrative data to study whether the lab and survey measures can predict which taxpayers evade taxes in the real world. Through this section, it is important to keep in mind that these are just correlations and, as such, they should not be interpreted as causal effects.

We start by analyzing some selected outcomes that we believe ex-ante to be the most relevant for the decision to evade taxes. Arguably, the most important candidate for predicting tax evasion due to its central role in the literature is the survey measure of tax morale.

Figure 7.a shows the link between this survey measure (x-axis) and the share of taxpayers who under-reported their wages relative to third-party reports in 2016 (y-axis). Among the individuals who say that it is never OK to evade taxes, 13.3% under-reported their wages. In comparison, among individuals who say that it is sometimes or always OK to evade taxes, 14.0% under-reported their wages. The difference between these two evasion rates is small (0.7 pp) and statistically insignificant (p-value=0.503). Most important, the difference is precisely estimated: according to the 95% CI, we can rule out that individuals with low tax morale evade on average over 2.69 pp more than individuals with high tax morale.

Next, we turn to the lab-based measure of tax morale, based on the game on evasion aversion. Figure 7.b splits the sample by respondents with low, medium, and high aversion to tax evaders. We find that the rates of wage under-reporting are similar across the three groups (13.3%, 13.4%, and 13.7%, respectively), and these differences are precisely estimated and statistically insignificant (p-value=0.942).

Figure 7.c corresponds to contributions in the public good game. We split the sample into individuals with low contributions to the public good (0–25% of their endowment), middle contributions (50%), and high contributions (75–100%). We find very small differences between the shares of wage under-reporters across the three groups: 13.6%, 13.1%, and 13.9%, respectively. These differences are precisely estimated and statistically insignificant (p-value=0.704).

Figure 7.d presents the results for the lab measure of dishonesty based on the dice game. This figure shows the share of wage under-reporters for groups of individuals who report each number from 1 through 6. Recall that the liars are concentrated among individuals who report 3, 4, or 5. As a result, we expect higher shares of wage under-reporters among those respondents. Instead, we find that the share of wage under-reporters is similar among the individuals who report $1, 2, 6$ (13.3%) and those who report $3, 4, 5$ (13.7%), and the difference is statistically insignificant (p-value=0.682). For a more direct comparison, we estimate a regression of whether the individual under-reported wages on the probability that he or she lied in the dice game, conditional on the number they reported (defined in Section 4.5). According to this regression, a 1 pp increase in the probability of lying in the dice game is associated with a decrease (rather than an expected increase) in the probability of under-reporting wages of just 0.07 pp, which is both economically small and statistically insignificant (p-value=0.981).

The first four panels from Figure 7 show that there is no relationship between the lab and survey measures and wage under-reporting. The last two panels show some significant correlations. Figure 7.e presents the results for the lab measure of generosity. One hypothesis is that individuals who are more altruistic will evade less because they care about the con-

sequences of tax evasion on the wellbeing of others (i.e., lower provision of public services). Alternatively, more altruistic individuals may believe that they could use those funds to help others better than the state would, and thus, will evade more. Consistent with the alternative hypothesis, we find that the more the individuals share in the dictator game, the more likely they are to be wage under-reporters: individuals who share nothing under-report wages at a rate of 11.5%, those who share a quarter of the endowment under-report wages at a rate of 12.8%, and those who share half or more of their endowment under-report wages at a rate of 14.71%. Moreover, we can reject the null hypothesis that the rates of wage under-reporting are the same across those three groups (p-value=0.006). While speculative, one potential explanation for the positive association between generosity and wage under-reporting could be that individuals who evade taxes feel guilty and thus want to compensate by being more generous. Indeed, some evidence along these lines has been reported in laboratory settings (Gneezy et al., 2014).[45]

Figure 7.f presents the results for the survey measure of the perceived evasion behavior of other employees. The hypothesis is that employees who think that other employees under-report their wages will find it "easier" to evade taxes themselves, presumably because they feel they are not breaking a social norm. We find some support for that hypothesis. Figure 7.f shows that the rate of wage under-reporting increases monotonically with the belief about the evasion behavior of others. The slope of that relationship indicates that for each 1 pp increase in the perceived share of wage under-reporters, the individual's own likelihood of wage under-reporting increases by 0.07 pp (p-value=0.003). This effect is small but not negligible.

5.2 Pairwise Correlations

The previous analysis focuses on a selected set of measures that are ex-ante expected to be most related to tax evasion. Here, we extend the analysis to all 24 lab and survey measures.

Table 2 reports the pairwise correlations between the tax evasion outcome (i.e., a dummy variable that equals 1 if the individual under-reported wages in 2016) and each of the 24 lab and survey measures (column (2)), jointly with the corresponding confidence intervals (column (2)) and p-values (column (3)). Because we are evaluating the significance of 24 different correlations, a natural concern is false positives due to multiple hypothesis testing. To provide an accurate assessment of the statistical significance of each correlation, while column (3) reports the p-values, column (4) reports the corresponding q-values to account

[45] An alternative interpretation is that some personality traits (e.g., political ideology) are positively correlated to generosity. This interpretation is consistent with the evidence presented below that measures related to political ideology have some ability to predict evasion choices.

for multiple hypothesis testing (Benjamini and Yekutieli, 2001).

Most correlation coefficients from Table 2 are economically small and precisely estimated around zero. The absolute value of the correlation coefficients ranges from 0 to 0.067. Moreover, due to the large sample sizes, each coefficient is precisely estimated, and the 95% CI indicates that we can often rule out even small correlations. For example, in the case of tax morale, the point estimate for the correlation is close to zero (-0.006) and the 95% CI ranges from -0.032 to 0.019. In other words, we can rule out even small correlations between the tax morale measure and the actual tax evasion decision. Indeed, in all 24 cases, the 95% confidence intervals rule out correlations above 0.10 (in absolute value), so we can confidently reject moderate to large correlations.

Among the lab measures, the highest correlations are given by the three lab measures related to generosity: correlations of 0.040 (p-value=0.002, q-value=0.020), 0.046 (p-value<0.001, q-value<0.001), and 0.048 (p-value<0.001, q-value<0.001) for the decisions to give to strangers, a charity, and the government, respectively. The survey measures are slightly more correlated with the evasion choices. Seven out of the 12 survey measures have correlations that are statistically significant after accounting for multiple-hypothesis testing (defined as q-values below 0.1), although all correlations tend to be small in magnitude. The strongest correlation is the perception of evasion of firms, with a correlation coefficient of 0.067 (p-value<0.001, q-value<0.001). The second-highest correlation is with respect to the left-right political spectrum, with a correlation coefficient of -0.062 (p-value<0.001, q-value<0.001). The third-highest correlation is with respect to the desired progressiveness, with a correlation coefficient of 0.058 (p-value<0.001, q-value<0.001). These correlations seem to indicate that the "type" of individuals who are more likely to evade are those who think that firms evade taxes, who are politically left-leaning, and who think that taxes should be more progressive.

In Appendix A.4, we show that these results are similar if we use alternative measures of tax evasion (e.g., deduction over-reporting instead of wage under-reporting or intensive margin instead of extensive margin) and sample definition (e.g., by restricting the sample to individuals with only one job).

5.3 Accounting for Measurement Error

One potential concern with the previous results is that, due to measurement error in the lab measures, the correlation coefficients may be subject to attenuation bias.[46] Perhaps after accounting for attenuation bias the correlations would be economically more important. We use two standard strategies to show that measurement error is unlikely to change the main

[46] A related concern is that of temporal stability. However, previous studies suggest that these measures can be relatively stable (Chuang and Schechter, 2015; Stango and Zinman, 2020)

conclusions.

The first and main strategy, described in Gillen et al. (2019), consists of using a correction factor for the attenuation bias based on the correlation between two elicitations of the same lab-measure (also known as a test-retest reliability). Because in our survey we do not elicit the same lab-measure twice, we use related studies to obtain a range of potential values of the scaling factor. Table 3 summarizes the results of this strategy. Column (1) corresponds to the correlation coefficient between each of the 12 lab measures and the measure of tax evasion (reproduced from column (1) of Table 2). Column (2) presents the range of the scaling factor reported in other studies (with column (3) providing the data source). While reports the re-scaled bounds corresponding to each correlation coefficient. For 9 our of the 12 measures, including the main measures such as *Evasion Aversion*, *Public Goods* and *Dishonesty*, the rescaled factors are still close to zero even after accounting for attenuation bias. However, for 3 out of the 12 measures (*Giving to charity, Giving to government,* and *Ultimatum*), the upper bounds for the rescaled correlations are quite high, meaning that we should take those null results with a grain of salt.

The second strategy that we implement, also discussed in Gillen et al. (2019), is the Obviously Related Instrumental Variables model. Specifically, we treat our lab measures as endogenous variables and some closely-related survey measures as their corresponding instrumental variable. For example, we instrument the lab measure *Evasion Aversion* using the survey measure *Tax Morale*. Due to the availability of variables that could be plausibly used as obviously related instruments, we were able to carry out this strategy for 3 of the 12 lab measures. The results are presented in Table 4. Each row corresponds to a different regression with a single right-hand-side variable, normalized to have a mean of 0 and a standard deviation of 1. The dependent variable is always an indicator variable for whether the taxpayer under-reported wages in 2016. Columns (1) and (2) presents the endogenous variable and the instrument(s) used in each case. As a benchmark for what the results look like without the measurement error correction, column (3) shows the OLS coefficients, which are related to, but should not be confused with, the pairwise correlations presented in column (1) of Table 2. Column (4) shows the 2SLS coefficients, which can be directly compared to the corresponding OLS coefficients from column (3). And to diagnose weak instruments, we report the Stock and Yogo (2005) F-statistic – in all four regression we can confidently reject the null hypothesis of weak instruments. For the most important measure, *Evasion Aversion*, we again find that the results are unchanged after accounting for measurement error. More precisely, both the OLS and 2SLS estimates of the slope (0.008 and -0.017, respectively) are close to zero and also close to each other. For *Giving to government* the coefficient remains similar after accounting for measurement error (0.048 in OLS vs 0.105 in 2SLS). For *Trust*,

the 2SLS slope (0.164) is a bit more different than the OLS counterpart (-0.008), meaning that the 2SLS correction is more substantial. However, even in that case the 2SLS slope is statistically insignificant.

5.4 Predictive Analysis

It is possible that none of the lab and survey measures have much predictive power on their own but have substantial joint explanatory power. To explore this possibility, Table 5 presents the results from a multivariate Probit regression. Each column corresponds to a different regression where the dependent variable is always an indicator variable that equals 1 for individuals who, according to third-party reports, under-reported their wages in 2016. At the bottom of the table, we present the traditional measure of predictive power with binary dependent variables: the area under the receiver operating characteristic (AUC). Assume that we randomly select an individual who evaded and another who did not and that we ask the model to identify the evader. The AUC tells us the probability that the model would guess correctly. A value of 0.5 indicates that the model does no better than chance (i.e., the same probability as flipping a coin). The greater the AUC, the better the predictive power, with a maximum value of 1 corresponding to 100% accuracy. For the sake of completeness, the table also reports an alternative measure of goodness-of-fit: the pseudo-R^2.

In column (1) of Table 5, the independent variables are the 12 lab measures (for their definitions, see the notes to Table 2). To make the coefficients more directly comparable to each other, the 24 lab and survey measures used as independent variables are normalized to have a mean of 0 and a standard deviation of 1. Only three coefficients are statistically significant, and they correspond to the same three variables that have the highest pairwise correlations: giving to strangers (p-value=0.056), giving to charity (p-value=0.002), and giving to the government (p-value<0.001). The AUC (0.564) indicates that the lab measures do slightly better than chance at identifying the tax evader.[47] The conclusions are similar for the alternative measure of goodness-of-fit, the pseudo-R^2 (0.008).

In column (2) of Table 5, the independent variables are the 12 survey measures (for their definitions, see the notes to Table 2). Similar to the lab measures, these 12 measures are standardized. Only three out of 12 have statistically significant coefficients, and they correspond to the same three variables with the strongest pairwise correlations: perceived evasion rate among firms (p-value<0.001), left-right political spectrum (p-value=0.046), and desired progressiveness (p-value=0.032). The predictive power of the survey measures (AUC=0.587, from column (2)) is a bit larger than the corresponding values for the lab

[47] These and other results are similar if we use the out-of-sample definition of the AUC instead. For example, the AUC in column (1) is 0.564 while the corresponding out-of-sample equivalent is 0.546.

measures (AUC=0.564, from column (1)) but far from perfect. In column (3) of Table 5, the independent variables are the combination of the 24 lab and survey measures. Again, the predictive power is a bit better (AUC=0.602, from column (3)) but still rather small.

The evidence so far suggests that, even when taken all at once, the measures of values and beliefs do a bad job at predicting which taxpayers evade taxes. However, it is often challenging to interpret the magnitudes of goodness of fit. For example, a low goodness of fit could imply that the outcome is completely random and thus impossible to predict with *any* type of data, not just these survey and lab measures. Thus, columns (4) through (6) of Table 5 provide results under three alternative models that can serve as benchmarks. The first benchmark is past behavior. Column (4) corresponds to a regression where the independent variables are a set of dummies about past evasion behavior: whether the taxpayer under-reported wages in 2015, over-reported deductions in 2015, or over-reported withholdings in 2015.[48] This regression is intended to identify persistence in the evasion choices. This persistence may reflect a host of time-invariant factors including but not limited to individual traits such as honesty.

The second benchmark tries to capture the standard economic incentives behind tax evasion choices. First, the workhorse model of tax evasion predicts that individuals will evade more when they face higher marginal tax rates (Allingham and Sandmo, 1972). Due to the tax schedule for personal income taxes in Uruguay in 2016, subjects can face one of 7 different marginal tax rates. To capture this mechanism, column (5) of Table 5 corresponds to a regression where the independent variables are a set of dummies for the 7 different marginal tax rates the individual faces according to the level of the third-party reported salary. And since the income level may also be an influential factor for cost-benefit analysis behind evasion choices (Allingham and Sandmo, 1972), the specification in column (5) also includes as independent variable the third-party reported wage level (more precisely, as a set of decile dummies).[49]

The third and last benchmark is the employee's context. Some individuals may find it easier to evade due to institutional factors. For example, if a firm makes salary information easier to access or more salient, it may be more difficult for its employees to under-report without feeling guilty. Moreover, the individual's context may matter due to social learning: employees may get tips on the opportunity to evade from their coworkers or even the firm's accountants. To proxy for these contextual factors, column (6) of Table 5 uses a regression

[48] A total of 191 taxpayers in the sample cannot be categorized as wage under-reporters (or deduction or withholding under-reporters) in 2015 because they did not file a tax return in 2015 or because they were not wage earners in 2015. To adjust for this, we control for a dummy variable that takes the value 1 for these 191 taxpayers and 0 otherwise.

[49] In the model of Allingham and Sandmo (1972), the sign of the expected effect of "real" income on declared income is ambiguous, depending for example on the assumptions about the shape of the utility function.

in which the independent variables correspond to the evasion choices of coworkers: i.e., the share of coworkers who under-reported wages in 2016, the share who over-reported deductions in 2016, and the share who over-reported withholdings in 2016.[50]

The results from Table 5 show that, on the one hand, the model that uses all 24 lab and survey variables at once has a bit higher predictive power (AUC=0.602, from columns (3)) than the model that uses past behavior (AUC=0.575, from column (4)). However, even with all 24 lab and survey variables at once, that model does worse than the other two benchmarks. The simple model of economic incentives has higher predictive power (AUC=0.615, from column (5)). And the model based on peer behavior fares much better (AUC=0.872, from column (6)). That is, although individual beliefs and values have *some* power to predict who evades taxes, the economic incentives and the individual's context play a much bigger role.

Indeed, our results that the peer behavior is most predictive of evasion choices is consistent with a small but growing literature showing that contextual factors, such as social learning, are important determinants of tax evasion. For example, Chetty et al. (2013) show significant geographic variation in how firms take advantage of the Earned Income Tax Credit for tax evasion, which they attribute to differences in knowledge (see also Wilson, 2020). Exploiting quasi-random variation in audit policy, Battaglini et al. (2020) document a robust correlation between evasion among clients of the same tax practitioner and provide evidence that this is partially due to information spillovers. Drago et al. (2020) use a field experiment that varied the content of mailings sent to potential evaders of TV license fees. They document a strong treatment spillover: untreated households are more likely to switch from evasion to compliance in response to mailings received by their network neighbors. Boning et al. (2020) presents results from a field experiment involving in-person visits by revenue officers. They show that these visits have a large direct effect on visited firms' tax deposits and the tax deposits of the firms' clients, suggesting a network effect whereby tax preparers disseminate information.

6 Conclusions

In this paper, we study how social preferences (e.g., dishonesty) and perceived social norms (e.g., whether others evade taxes) make individuals more prone to evading taxes. We exploit administrative tax data on wage earners' tax returns and their employers' third-party reports to measure tax evasion of personal income tax in Uruguay. We provide evidence of

[50] These variables are undefined for 82 of the taxpayers who do not have coworkers in 2016. To account for this, we include as additional control variable a dummy that takes the value 1 for these 82 taxpayers and 0 otherwise.

significant variation in evasion choices across taxpayers: while some employees evade taxes by under-reporting their wages (or over-reporting deductions or withholdings), other employees report truthfully. In collaboration with Uruguay's tax agency, we conducted a survey with a subsample of 6,078 taxpayers to measure their preferences and beliefs using survey questions and laboratory games with real financial incentives. Surprisingly, some of the measures that should in theory be most related to tax evasion choices, such as the survey measure of tax morale, have no power whatsoever to identify tax evaders. Other measures have *some* predictive power. Most notably, taxpayers who are more likely to under-report wages tend to perceive firms as evaders, to identify as politically left-leaning, to think that taxes should be more progressive, and to be more generous. However, the 24 lab and survey measures, even when considered jointly, do not have as much predictive power as the economic incentives or the evasion behavior of coworkers.

Our findings suggest that either these individual traits are not relevant for evasion choices, or that lab experiments and survey questions cannot adequately identify meaningful heterogeneity in those individual traits.[51] On the other hand, even if lab measures are not good at predicting who evades more, they arguably do a decent job at predicting aggregate evasion behavior. In the dice game, we estimate that 11% of taxpayers lied about the roll of their dice. This finding is consistent in magnitude with the aggregate evasion rate, according to which 15.5% under-report wages.[52] Indeed, there is another aggregate pattern in the dice game that seem largely consistent with the distribution of wage under-reporting. In our dice game, even those who lie, they do not lie all the way (i.e., reporting the number 6) – indeed, this is a robust finding in the literature (Fischbacher and Föllmi-Heusi, 2013). In comparison, we find that taxpayers almost never fully under-report wages: conditional on under-reporting, the average rate of under-reporting is just 6.1%. Last, one robust finding from the literature on tax evasion is that taxpayers are less likely to lie in tax forms when the lie can be more easily detected ex-post, such as when there are third-party reports available (Kleven et al., 2011; Pomeranz, 2015; Naritomi, 2019). This same result is present in our laboratory experiments: taxpayers lied in the dice game, where the lie is not ex-post verifiable, but virtually never lied about their date of birth, which was ex-post verifiable.

[51] One possibility is that individual traits, such as dishonesty or altruism, vary across individuals but those differences are not picked up by the lab and survey measures. Alternatively, perhaps these traits vary little across individuals and thus most of the variation in lab and survey measures is just measurement error.

[52] In a similar spirit, Alm et al. (2015) compare the distribution of tax evasion in a laboratory setting and shows that it is similar, in aggregate, to the distribution of tax evasion in the real world.

References

Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for Truth-Telling. *Econometrica 87*(4), 1115–1153.

Allingham, M. G. and A. Sandmo (1972). Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics 1*(3-4), 323–338.

Alm, J., K. M. Bloomquist, and M. McKee (2015). On the External Validity of Laboratory Tax Compliance Experiments. *Economic Inquiry 53*(2), 1170–1186.

Alm, J., B. Jackson, and M. McKee (1992). Estimating the Determinants of Taxpayer Compliance with Experimental Data. *National Tax Journal 45*(1), 107–114.

Alm, J., E. Kirchler, and S. Muehlbacher (2012). Combining Psychology and Economics in the Analysis of Compliance: From Enforcement to Cooperation. *Economic Analysis and Policy 42*(2), 133–151.

Alm, J., I. Sanchez, and A. De Juan (1995). Economic and Noneconomic Factors in Tax Compliance. *Kyklos 48*(1), 1–18.

Almås, I., A. W. Cappelen, and B. Tungodden (2020). Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians? *Journal of Political Economy 128*(5), 1753–1788.

Andreoni, J., B. Erard, and J. Feinstein (1998). Tax Compliance. *Journal of Economic Literature 32*(2), 818–860.

Battaglini, M., L. Guiso, C. Lacava, and E. Patacchini (2020). Tax Professionals and Tax Evasion. *NBER Working Paper No. 25745*.

Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics 29*(4), 1165–1188.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior 10*(1), 122–142.

Bergolo, M., G. Burdín, M. de Rosa, M. Giaccobasso, and M. Leites (2019). Tax Bunching at the Kink in the Presence of Low Capacity of Enforcement: Evidence From Uruguay. *IZA Discussion Paper No. 12286*.

Bergolo, M. L., R. Ceni, G. Cruces, M. Giaccobasso, and R. Perez-Truglia (2021). Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment. *American Economic Journal: Economic Policy, forthcoming*.

Best, M. C. (2014). The Role of Firms in Workers' Earnings Responses to Taxes: Evidence From Pakistan. *Working Paper*.

Biro, A., D. Prinz, and L. Sandor (2020). Tax Evasion and the Minimum Wage. *Working Paper*.

Bjørneby, M., A. Alstadsæter, and K. Telle (2021). Limits to third-party reporting: Evidence from

a randomized field experiment in Norway. *Journal of Public Economics 203*, 104512.

Blumenthal, M., C. Christian, and J. Slemrod (2001). Do Normative Appeals Affect Tax Compliance? Evidence From a Controlled Experiment in Minnesota. *National Tax Journal 54*(1), 125–138.

Boning, W. C., J. Guyton, R. Hodge, and J. Slemrod (2020). Heard it through the grapevine: The direct and network effects of a tax enforcement field experiment on firms. *Journal of Public Economics 190*, 104261.

Bott, K. M., A. W. Cappelen, E. Ã. Sørensen, and B. Tungodden (2020). Youâve Got Mail: A Randomized Field Experiment on Tax Evasion. *Management Science 66*(7), 2801–2819.

Bottan, N. L. and R. Perez-Truglia (2020). Choosing Your Pond: Location Choices and Relative Income. *The Review of Economics and Statistics, forthcoming*.

Brockmeyer, A. and M. Hernandez (2016). Taxation, information, and withholding: evidence from Costa Rica. *World Bank Policy Research Working Paper No. 7600*.

Brosig, J., T. Riechmann, and J. Weimann (2007). Selfish in the end?: An investigation of consistency and stability of individual behavior. *MPRA Munich Personal RePEc Archive* (2035).

Camerer, C. F. (2011). The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List. *SSRN Working Paper*.

Cappelen, A. W., J. Mollerstrom, B.-A. Reme, and B. Tungodden (2019). A Meritocratic Origin of Egalitarian Behavior. *Working Paper*.

Cardenas, J. C. and J. Carpenter (2008). Behavioural Development Economics: Lessons from Field Labs in the Developing World. *The Journal of Development Studies 44*(3), 311–338.

Carlsson, F., O. Johansson-Stenman, and P. K. Nam (2014). Social preferences are stable over long periods of time. *Journal of Public Economics 117*.

Carrillo, P., D. Pomeranz, and M. Singhal (2017). Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement. *American Economic Journal: Applied Economics 9*(2), 144–164.

Carson, R. T. and T. Groves (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics 37*(1), 181–210.

Castañeda, N., D. Doyle, and C. Schwartz (2020). Opting Out of the Social Contract: Tax Morale and Evasion. *Comparative Political Studies 53*(7), 1175–1219.

Castro, L. and C. Scartascini (2015). Tax Compliance and Enforcement in the Pampas Evidence From a Field Experiment. *Journal of Economic Behavior & Organization 116*, 65–82.

Charness, G., U. Gneezy, and B. Halladay (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization 131*, 141–150.

Charness, G. and M. Rabin (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics 117*(3).

Chetty, R., J. N. Friedman, and E. Saez (2013). Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings. *The American Economic Review 103*(7), 2683–2721.

Chuang, Y. and L. Schechter (2015). Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics 117*, 151–170.

Cohn, A. and M. A. Maréchal (2018). Laboratory Measure of Cheating Predicts School Misconduct. *The Economic Journal 128*(615), 2743–2754.

Coller, M. and M. B. Williams (1999). Eliciting Individual Discount Rates. *Experimental Economics 2*(2), 107–127.

Cowell, F. A. and J. P. F. Gordon (1988). Unwillingness to pay: Tax evasion and public good provision. *Journal of Public Economics 36*(3), 305–321.

Crede, A.-K. and F. von Bieberstein (2020). Reputation and lying aversion in the die roll paradigm: Reducing ambiguity fosters honest behavior. *Managerial and Decision Economics 41*(4), 651–657.

Cummings, R. G., J. Martinez-Vazquez, and M. McKee (2006). Cross Cultural Comparisions of Tax Compliance Behavior. *Journal of Economic Psychology 27*(2), 224–246.

Dai, Z., F. Galeotti, and M. C. Villeval (2018). Cheating in the Lab Predicts Fraud in the Field: An Experiment in Public Transportation. *Management Science 64*(3), 1081–1100.

Dean, M. and A. Sautmann (2021). Credit constraints and the measurement of time preferences. *Review of Economics and Statistics 103*(1).

DeBacker, J., B. T. Heim, and A. Tran (2015). Importing corruption culture from overseas: Evidence from corporate tax evasion in the United States. *Journal of Financial Economics 117*(1), 122–138.

Dellavigna, S. and E. Linos (2021). RCTs to Scale: Comprehensive Evidence from Two Nudge Units. *Econometrica, forthcoming*.

Doerrenberg, P. and A. Peichl (2013). Progressive taxation and tax morale. *Public Choice 155*(3-4), 293–316.

Doerrenberg, P. and A. Peichl (2018). Tax Morale and the Role of Social Norms and Reciprocity. Evidence from a Randomized Survey Experiment. *CESifo Working Paper No. 7149*.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010). Are Risk Aversion and Impatience Related to Cognitive Ability? *American Economic Review 100*(3), 1238–1260.

Drago, F., F. Mengel, and C. Traxler (2020). Compliance Behavior in Networks: Evidence from a Field Experiment. *American Economic Journal: Applied Economics 12*(2), 96–133.

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics 14*, 583–610.

Epper, T., E. Fehr, H. Fehr-Duda, C. T. Kreiner, D. D. Lassen, S. Leth-Petersen, and G. N. Rasmussen (2020). Time Discounting and Wealth Inequality. *American Economic Review 110*(4), 1177–1205.

Exley, C. and J. Kessler (2019). Motivated Errors. *NBER Working Paper No. 26595*.

Falk, A., A. Becker, T. Dohmen, D. B. Huffman, and U. Sunde (2018). Global Evidence on Economic Preferences. *The Quarterly Journal of Economics 133*(4), 1645–1692.

Fehr, E. (2009). On The Economics and Biology of Trust. *Journal of the European Economic Association 7*(2-3), 235–266.

Feld, L. P. and B. S. Frey (2002). Trust breeds trust: How taxpayers are treated. *Economics of Governance 3*(2), 87–99.

Fellner, G., R. Sausgruber, and C. Traxler (2013). Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information. *Journal of the European Economic Association 11*(3), 634–660.

Fischbacher, U. and F. Föllmi-Heusi (2013). Lies in Disguise - An Experimental Study on Cheating. *Journal of the European Economic Association 11*(3), 525–547.

Frey, B. S. and B. Torgler (2007). Tax morale and conditional cooperation. *Journal of Comparative Economics 35*(1), 136–159.

Gächter, S. and J. F. Schulz (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature 531*(7595), 496–499.

Gillen, B., E. Snowberg, and L. Yariv (2019). Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study. *Journal of Political Economy 127*(4), 1826–1863.

Gneezy, U., A. Imas, and K. Madarász (2014). Conscience Accounting: Emotion Dynamics and Social Behavior. *Management Science 60*(11), 2645–2658.

Gomez-Sabaini, J. C. and J. P. Jimenez (2012). Tax structure and tax evasion in Latin America. *Macroeconomics of Development Series 118*.

Gomez-Sabaini, J. C. and D. Moran (2014). Tax policy in Latin America Assessment and guidelines for a second generation of reforms. *Macroeconomics of Development Series 133*.

Halla, M. (2012). Tax Morale and Compliance Behavior: First Evidence on a Causal Link. *The B.E. Journal of Economic Analysis & Policy 12*(1), 1–27.

Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics 148*, 14–31.

Hanna, R. and S.-Y. Wang (2017). Dishonesty and Selection into Public Service: Evidence from India. *American Economic Journal: Economic Policy 9*(3), 262–290.

Hvidberg, K. B., C. Kreiner, and S. Stantcheva (2020). Social Position and Fairness Views. *NBER Working Paper No. 28099*.

Internal Revenue Service (2016). Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2008-2010. Technical report.

Jensen, A. (2019). Employment Structure and the Rise of the Modern Tax System. *NBER Working Paper No. 25502*.

Johns, A. and J. Slemrod (2010). The Distribution of Income Tax Noncompliance. *National Tax Journal 63*(3), 397–418.

Karadja, M., J. Mollerstrom, and D. Seim (2017). Richer (and Holier) Than Thou? The Effect of Relative Income Improvements on Demand for Redistribution. *The Review of Economics and Statistics 99*(2), 201–212.

Kleven, H. J. (2014). How Can Scandinavians Tax So Much? *Journal of Economic Perspectives 28*(4), 77–98.

Kleven, H. J., M. B. Knudsen, T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or Unable to Cheat? Evidence from a Randomized Tax Audit Experiment in Denmark. *Econometrica 79*(3), 651–692.

Kleven, H. J., C. Kreiner, and E. Saez (2016). Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries. *Economica 83*, 2016.

Kreiner, C. T., D. Dreyer Lassen, and S. Leth-Petersen (2019). Liquidity Constraint Tightness and Consumer Responses to Fiscal Stimulus Policy. *American Economic Journal: Economic Policy 11*(1), 351–379.

Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015). How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments. *American Economic Review 105*(4), 1478–1508.

Levin, I. P., S. S. Hart, J. A. Weller, and L. A. Harshman (2007). Stability of choices in a risky decision-making task: A 3-year longitudinal study with children and adults. *Journal of Behavioral Decision Making 20*(3).

Levitt, S. D. and J. A. List (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives 21*(2), 153–174.

Li, S. X., C. C. Eckel, P. J. Grossman, and T. L. Brown (2011). Giving to government: Voluntary taxation in the lab. *Journal of Public Economics 95*(9-10), 1190–1201.

List, J. A. (2004). Young, Selfish and Male: Field Evidence of Social Preferences. *The Economic Journal 114*(492), 121–149.

Lönnqvist, J. E., M. Verkasalo, G. Walkowitz, and P. C. Wichardt (2015). Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison. *Journal of Economic Behavior and Organization 119*.

Luttmer, E. F. P. and M. Singhal (2014). Tax Morale. *Journal of Economic Perspectives 28*(4), 149–168.

Meier, S. and C. D. Sprenger (2015). Temporal stability of time preferences. *Review of Economics and Statistics 97*(2).

Naritomi, J. (2019). Consumers as Tax Auditors. *American Economic Review 109*(9), 3031–3072.

Nathan, B., R. Perez-Truglia, and A. Zentner (2020). My Taxes are Too Darn High: Why Do Households Protest their Taxes? *NBER Working Paper No. 27816*.

Oosterbeek, H., R. Sloof, and G. van de Kuilen (2004). Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis. *Experimental Economics 7*(2), 171–188.

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

Ross, L. and R. E. Nisbett (2011). *The person and the situation: Perspectives of social psychology*. London: Pinter & Martin Ltd.

Slemrod, J. (2008). Does It Matter Who Writes the Check to the Government? The Economics of Tax Remittance. *National Tax Journal 61*.

Slemrod, J. (2019). Tax Compliance and Enforcement. *Journal of Economic Literature 57*(4), 904–954.

Slemrod, J., M. Blumenthal, and C. Christian (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. *Journal of Public Economics 79*(3), 455–483.

Slemrod, J., B. Collins, J. Hoopes, D. Reck, and M. Sebastiani (2017). Does credit-card information reporting improve small-business tax compliance? *Journal of Public Economics 149*, 1–19.

Stango, V. and J. Zinman (2020). Behavioral Biases are Temporally Stable. *NBER Working Paper No. 27860*.

Stock, J. H. and M. Yogo (2005). Testing for Weak Instruments in Linear IV Regression. In D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models*, pp. 80–108. Cambridge University Press.

Torgler, B. (2005). Tax morale and direct democracy. *European Journal of Political Economy 21*(2), 525–531.

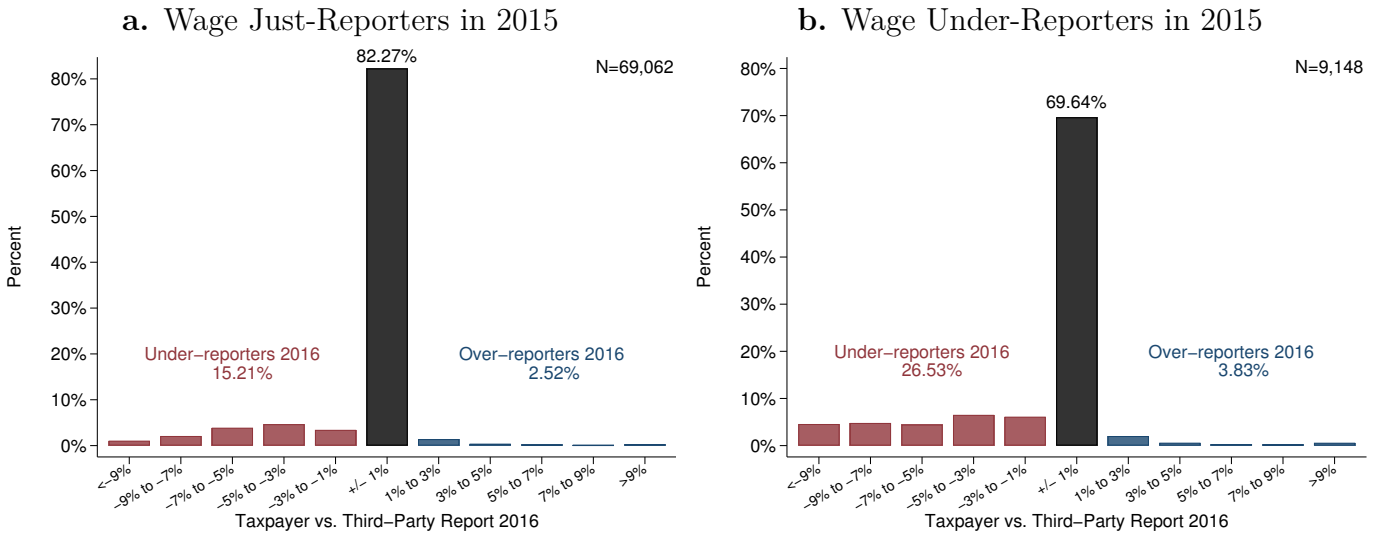
Wilson, R. (2020). The Impact of Social Networks on EITC Claiming Behavior. *Review of Economics and Statistics, forthcoming*.

World Values Survey Association (2014). World Values Survey Wave 6 (2010-2014) Official Aggregate v.20180912. Technical report, Asep/JDS, Madrid, Spain.

Figure 1: Measures of Tax Evasion

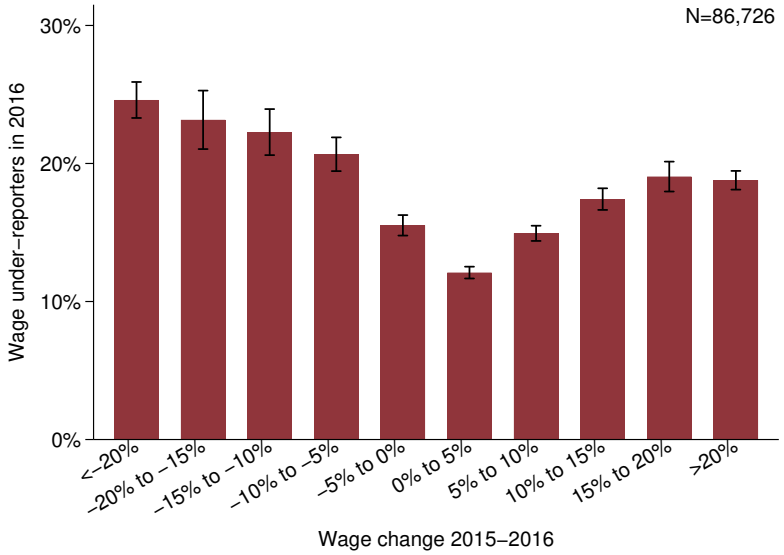
Notes: Panel (a): discrepancies between the wages reported in the taxpayer's tax return versus the employer's third-party report (as the percentage of third-party wage income). Results based on the sample of taxpayers who were pure wage earners and filed a tax return in 2016. Panel (b): discrepancies between the effective tax liability (net of deductions) of the taxpayers and the counterfactual tax liability they would face if they had reported wages equal to the third-party report (as a percentage of the latter). We must restrict the sample to taxpayers with positive counterfactual tax amounts (to avoid dividing by zero). Panel (c): discrepancies between the automatic deductions reported in the taxpayer's tax return versus the employer's third-party report (as a percentage of the latter). Results based on the subsample of taxpayers who reported their wages within 1% of their third-party reports (i.e., wage just-reporters). Panel (d): discrepancies between the tax withholdings reported in the taxpayer's tax return versus the employer's third-party report (as a percentage of the latter). Results based on the subsample of taxpayers who reported their wages within 1% of their third-party reports (i.e., wage just-reporters).

Figure 2: Persistence of Misreporting



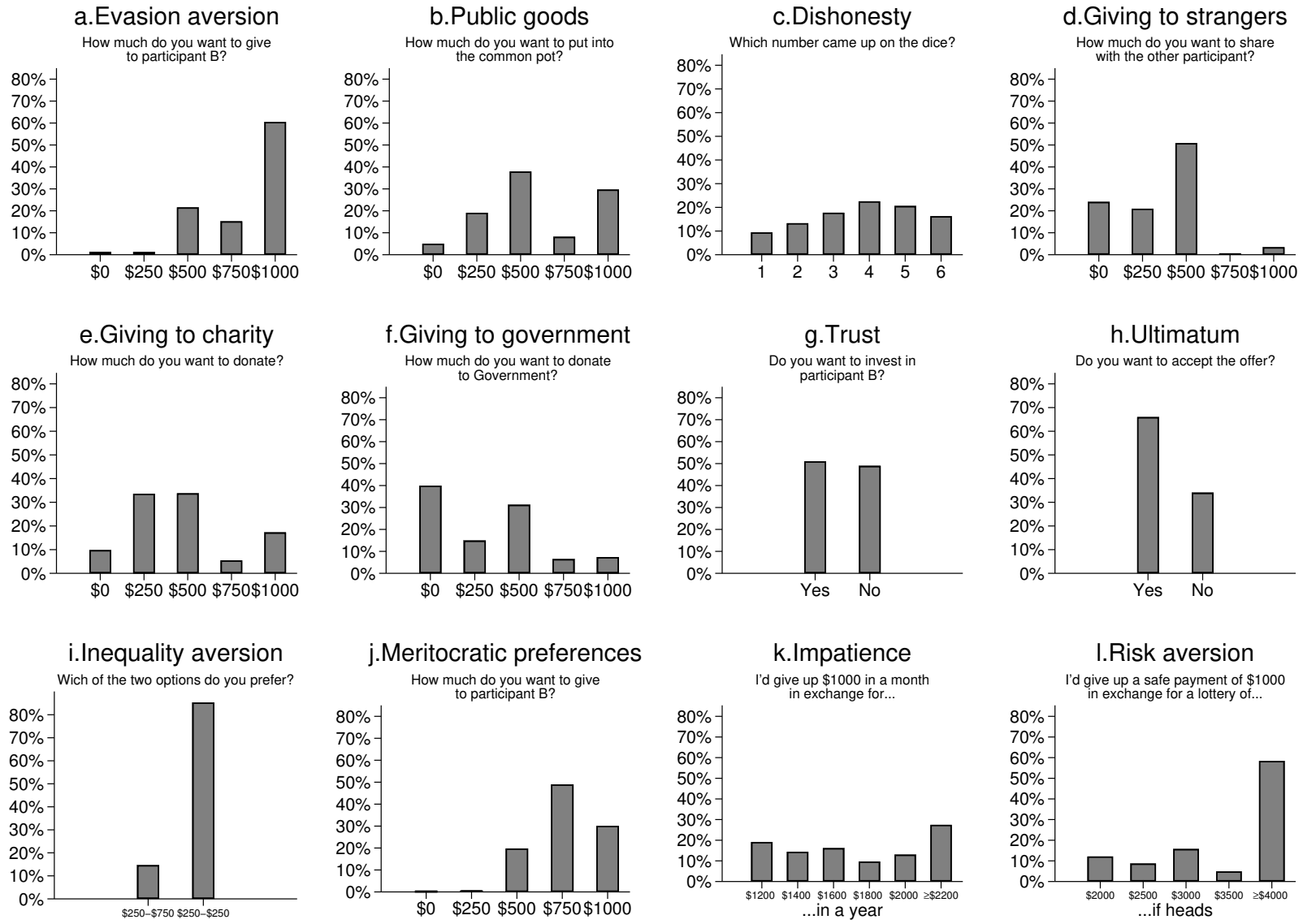
Notes: Discrepancies in wages reported in the taxpayer tax return versus the employer's third-party report. Results based on the sample of 86,749 taxpayers who were pure wage earners and filed a tax return in both 2015 and 2016. Panel (a) corresponds to the subsample of taxpayers who reported their wages within 1% of their third-party reports (i.e., just-reporters) in 2015, while panel (b) corresponds to the subsample of taxpayers who reported wages below 1% of their third-party reports (i.e., under-reporters) in 2015.

Figure 3: Wage Under-Reporting and Wage Changes



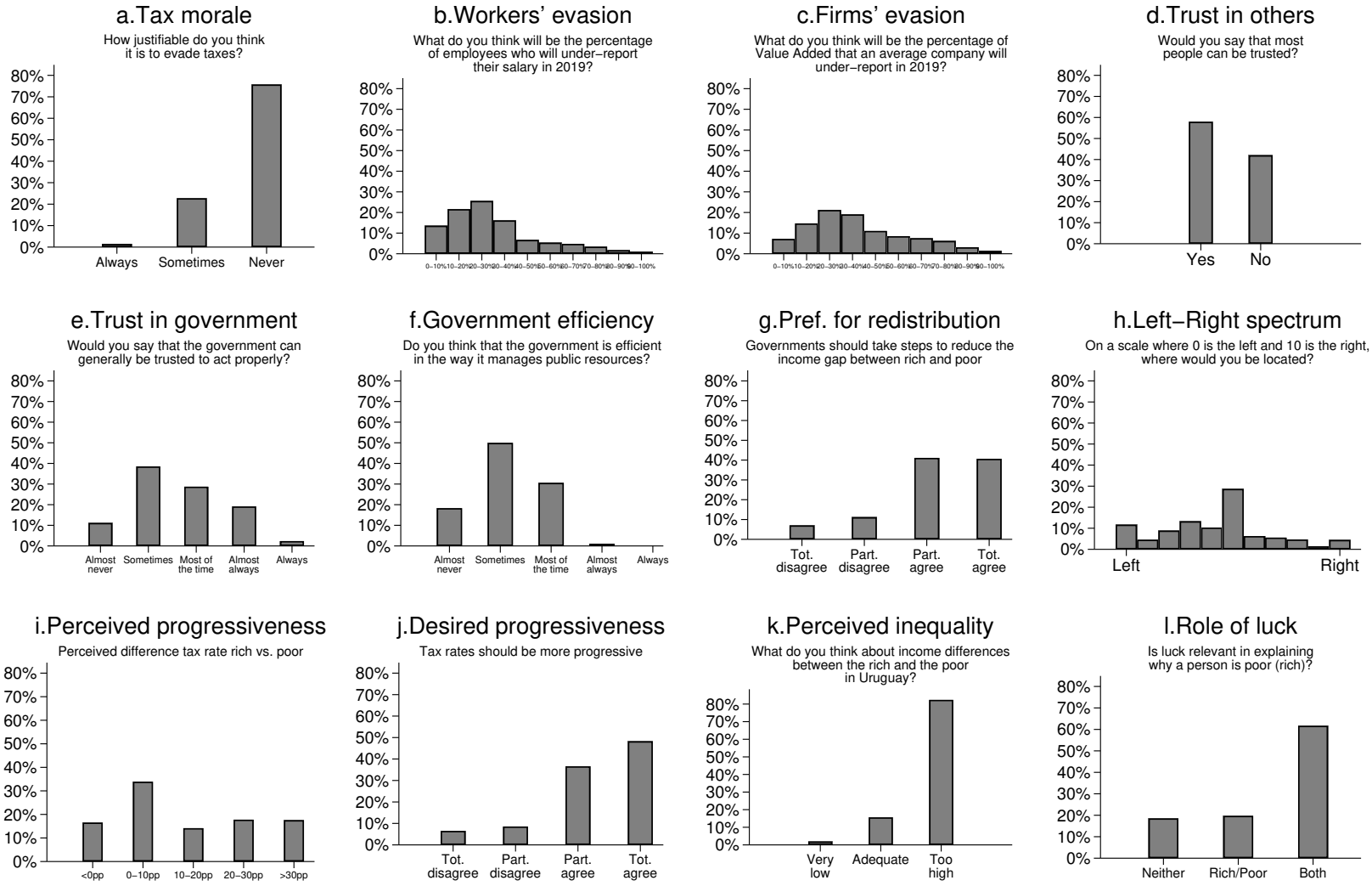
Notes: Each bar represents the share of employees under-reporting their wages in 2016 for a different group of employees. As denoted by the x-axis, the employees are grouped based on the change in their third-party reported wages from 2015 to 2016. These results are based on the sample of 86,726 taxpayers that are pure wage earners and file a tax return both in 2015 and 2016.

Figure 4: Distribution of Lab-Measures



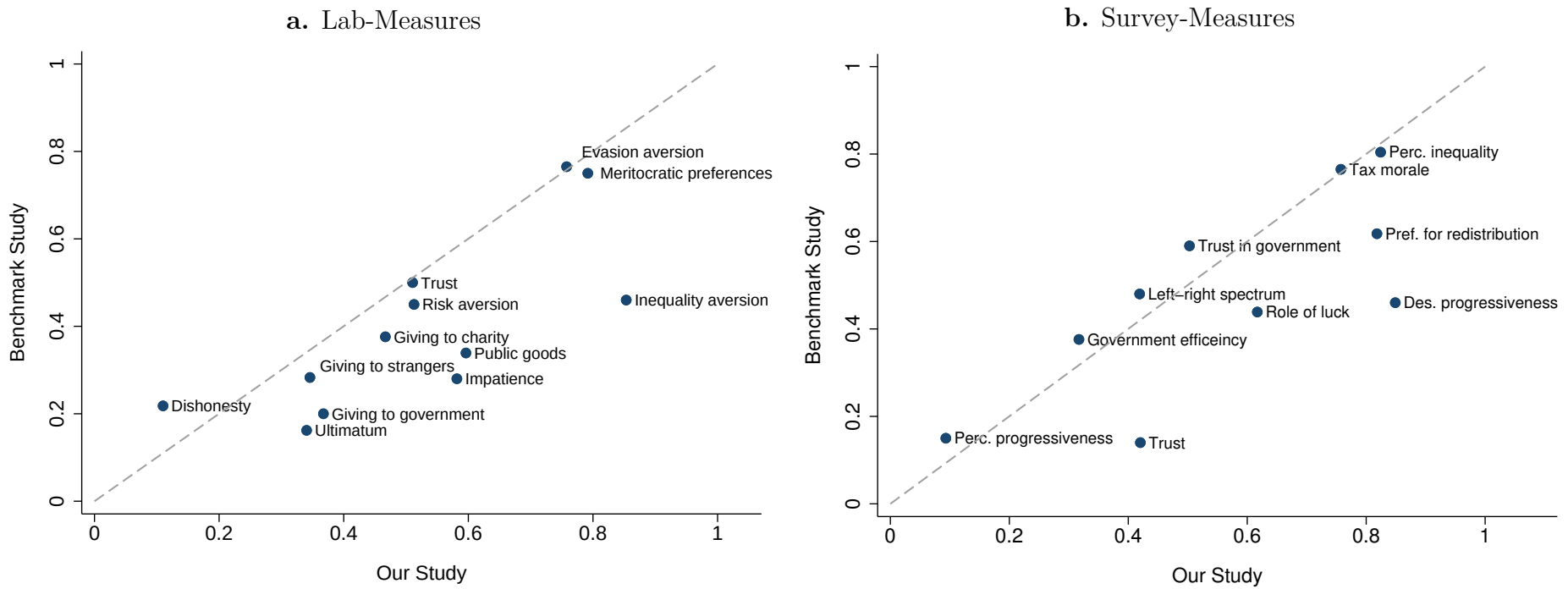
Notes: Distribution of the 12 lab-based measures among the 6,078 taxpayers who responded to the survey. For the full survey questionnaire, see Appendix C.

Figure 5: Distribution of Survey-Measures



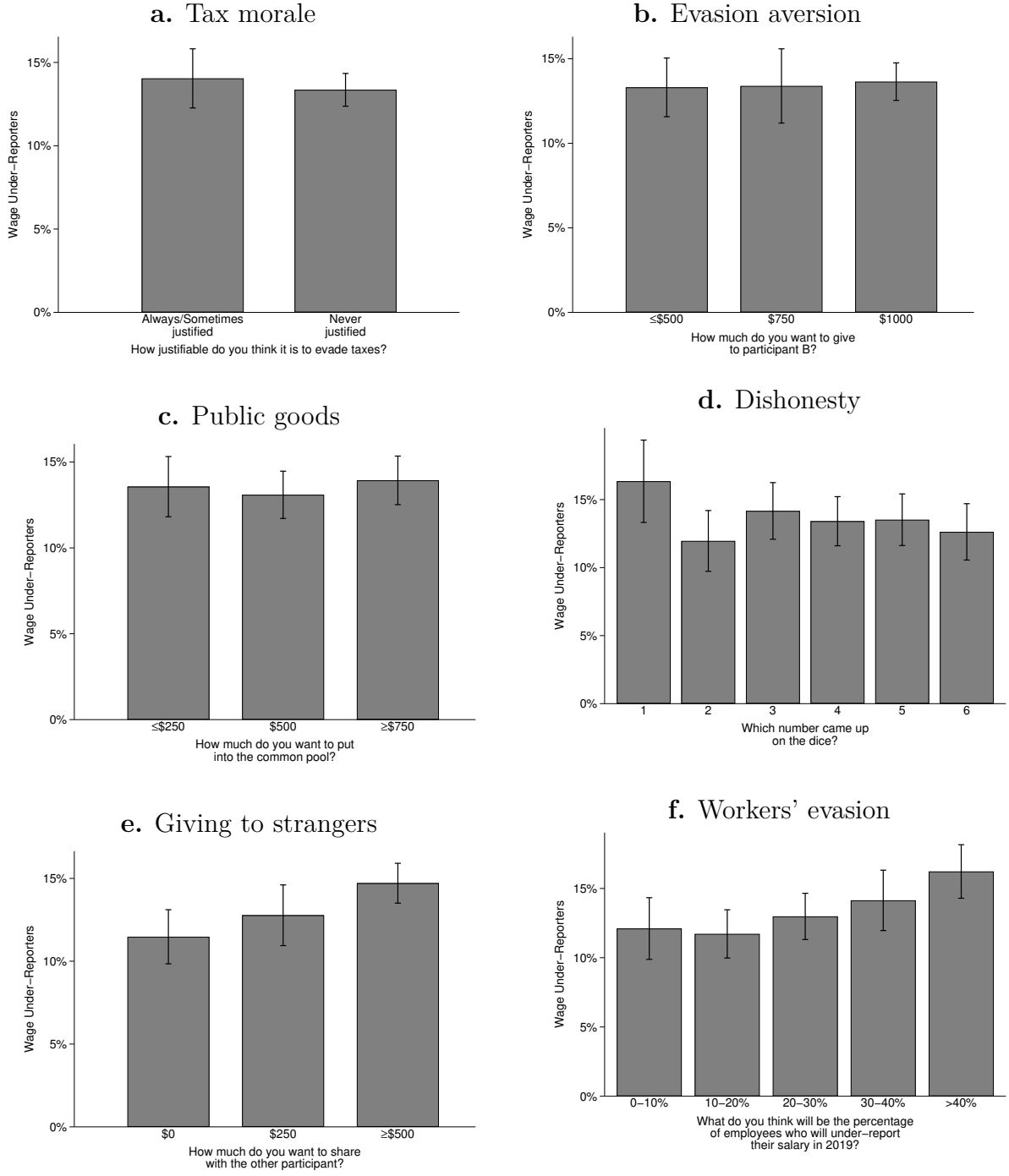
Notes: Distribution of the 12 survey-based measures among the 6,078 taxpayers who responded to the survey. For the full survey questionnaire, see Appendix C.

Figure 6: Comparison of the Results from Lab- and Survey-Measures to Related Studies



Notes: Comparison of average responses in our study (N=6,078) vs. the corresponding benchmark study. Lab-measures: *evasion aversion* is the share of respondents that give more endowment to those participants that think that evading taxes is never justifiable (benchmark study: Luttmer and Singhal (2014)), *public goods* is the average share of the endowment contributed to the common pool (List, 2004), *dishonesty* is the probability of lying, measured as the excess mass in faces 3, 4 and 5 (Gächter and Schulz, 2016), *giving to strangers* is the share of the endowment that the respondent gave to a randomly-chosen individual (Engel, 2011), *giving to charity* is the share of the endowment that the respondent gave to charity (Engel, 2011), *giving to government* is the share of the endowment that the respondent gave to the government ministry of social development (Li et al., 2011), *trust* is the share of respondents that decided to invest in the other participant (Cardenas and Carpenter, 2008), *ultimatum* is the share of respondents that rejected the offer (Oosterbeek et al., 2004), inequality aversion is the share of respondents that preferred an egalitarian allocation (Charness and Rabin, 2002), *Meritocratic preferences* is the share of respondents that gave a bigger share of the endowment to the participant that had made an effort (Cappelen et al., 2019), *impatience* is the extra-share of the payment that must be offered to the respondent in order for them to prefer waiting a year for the payment instead of perceiving the payment next month (Coller and Williams, 1999), *risk aversion* is the extra-share of the safe payment that must be offered to the respondent in order for them to prefer a 50-50 lottery instead of the safe payment (Dohmen et al., 2010). Survey-measures: *tax morale* is the share of respondents that thinks evading taxes is never justifiable (*), *trust in others* is the share of respondents that thinks most people can be trusted (*), *trust in government* is the share of respondents that thinks always, almost always or most of the time, can trust in the government acting properly (*), *government efficiency* is the share of respondents that thinks the government is efficient or very efficient (*), *preferences for redistribution* is the share of respondents that thinks the government should take steps to reduce the income gap between rich and poor (*), *Left-Right spectrum* is the average response to a question about where the individual would be located in a range in which 0 is left and 1 es right (**), *perceived progressiveness* is the respondents' perceived difference in tax rate paid between the upper 20% and the bottom 20% of the income distribution (***), *desired progressiveness* is the share of respondents that are totally or partially agree with making the tax system more progressive (**), *perceived inequality* is the share of respondents that thinks that inequality is too high (**), *role of luck* is the share of respondents that thinks that luck and other circumstances beyond individual control are more important than individual effort in becoming both rich and poor (*). The benchmarks for the survey measures are based on our own calculations with the following data sources: (*) World Values Survey Association (2014); (**) 2011 Latinobarometro (`https://www.latinobarometro.org/latContents.jsp`); (***) 2015/16 ELBU (`http://fcea.edu.uy/datos/bases-de-datos.html`).

Figure 7: Correlation between Actual Tax Evasion Choices and Selected Preferences and Beliefs



Notes: Results based on the 6,078 survey respondents. Each bar corresponds to the share of wage under-reporters (i.e., taxpayers who reported wage income below 1% of the third-party report filed by their employers), with 95% robust confidence intervals. Each panel breaks the same sample of 6,078 taxpayers into subgroups based on selected lab or survey measures: e.g., panel (a) breaks them down by the responses to the question on tax morale.

Table 1: Descriptive Statistics

	All	Tax filers	Pure WE	Invited		Responded	
				No	Yes	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female (%)	46.74	47.26	44.85	38.37	52.78	52.25	58.16
	(0.04)	(0.11)	(0.13)	(0.17)	(0.19)	(0.20)	(0.63)
Age	43.32	45.74	45.20	45.81	44.45	44.54	43.54
	(0.01)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.13)
Total income (USD)	20.78	39.57	37.86	34.54	41.90	41.96	41.36
	(0.04)	(0.11)	(0.12)	(0.17)	(0.17)	(0.18)	(0.40)
Wage earners (%)	96.58	86.89	100.00	100.00	100.00	100.00	100.00
	(0.02)	(0.08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Wage (USD)	19.44	39.15	36.45	33.03	40.64	40.66	40.41
	(0.03)	(0.09)	(0.09)	(0.11)	(0.15)	(0.16)	(0.38)
Wage Under-reporters (%)			15.49	16.90	13.76	13.78	13.52
			(0.09)	(0.13)	(0.13)	(0.14)	(0.44)
Autom. Deductions Over-reporters (%)			22.04	20.51	23.91	23.59	27.20
			(0.11)	(0.14)	(0.16)	(0.17)	(0.57)
Withholding Over-reporters (%)			22.88	18.56	28.19	28.16	28.52
			(0.11)	(0.14)	(0.17)	(0.18)	(0.59)
Observations	1,271,509	200,519	151,565	83,357	68,208	62,130	6,078

Notes: Average characteristics for the year 2016, with robust standard errors in parentheses. Column (1) corresponds to the whole unvierse of taxpayers. Column (2) corresponds to the group of taxpayers that file a tax return. Column (3) corresponds to the group of taxpayers that file a tax return and are pure wage earners (i.e do not receive self-employment income). Additionally, column (3) excludes those pure wage earners that have missing wage income data on their third-party report. Column (4) to column (7) correspond to sub groups of column (3). Column (4) and (5) correspond to the group of taxpayers that were not and were invited to the survey, respectively. Column (6) and (7) correspond to the group of taxpayers that did not respond and did respond to the survey, respectively. *Female* is the percentage of female. *Age* is the age average in years. *Total income* is the total annual income average in thousands of 2016 U$D. *Wage earners* is the percentage of wage earners. *Wage* is the average wage in thousands of 2016 U$D. For those that don't file a tax return, *Wage* is the third-party reported wage including bonus. For those that file a tax return, *Wage* is the self-reported wage including bonus income. *Wage Under-reporters* is the percentage of taxpayers that under-report their wage. *Autom. Deductions Over-reporters* is the percentage of taxpayers that over-report their automatic deductions. Finally, *Withholding Over-reporters* is the percentage of taxpayers that over-report their withholding income.

Table 2: Pairwise Correlations between Tax Evasion and Lab and Survey Measures

		Corr.Coef. (1)	95% CI (2)	p-value (3)	q-value (4)
Lab-Based	Evasion aversion	0.008	[-0.017,0.033]	0.541	0.999
	Public goods	0.003	[-0.022,0.028]	0.812	0.999
	Dishonesty	-0.000	[-0.025,0.025]	0.981	0.999
	Giving to strangers	0.040	[0.015,0.065]	0.002	0.020
	Giving to charity	0.046	[0.021,0.071]	<0.001	<0.001
	Giving to government	0.048	[0.023,0.073]	<0.001	<0.001
	Trust	-0.008	[-0.033,0.017]	0.516	0.999
	Ultimatum	0.026	[0.000,0.051]	0.046	0.321
	Inequality aversion	0.007	[-0.018,0.033]	0.560	0.999
	Meritocratic preferences	-0.001	[-0.026,0.024]	0.925	0.999
	Impatience	0.021	[-0.005,0.046]	0.108	0.652
	Risk aversion	0.013	[-0.012,0.038]	0.324	0.999
Survey-Based	Tax morale	-0.006	[-0.032,0.019]	0.613	0.999
	Workers' evasion	0.038	[0.013,0.064]	0.003	0.027
	Firms' evasion	0.067	[0.042,0.092]	<0.001	<0.001
	Trust in others	0.024	[-0.001,0.049]	0.063	0.408
	Trust in government	0.028	[0.003,0.054]	0.027	0.204
	Government efficiency	0.046	[0.021,0.071]	<0.001	<0.001
	Preferences for redistribution	0.042	[0.017,0.067]	0.001	0.011
	Left-right spectrum	-0.062	[-0.087,-0.037]	<0.001	<0.001
	Perceived progressiveness	-0.010	[-0.035,0.015]	0.445	0.999
	Desired progressiveness	0.058	[0.033,0.084]	<0.001	<0.001
	Perceived inequality	0.029	[0.003,0.054]	0.026	0.204
	Role of luck	0.047	[0.022,0.072]	<0.001	<0.001

Notes: Correlations between tax evasion (and an indicator variable that takes the value 1 if the individual under-reported wages in 2016 and 0 otherwise) and each of the the lab and survey measures. Results for the 6,078 wage earners who filed a tax return in 2016 and responded to our survey. The q-values (column (4)) are based on the Yekutieli method. *Evasion aversion* is the share of the endowment given to the partner who thinks evading taxes is never justifiable, *public goods* is the share of the endowment contributed to the common pool, *dishonesty* is the probability of having lied conditional on the dice draw that was reported, *giving to strangers* is the share of the endowment given to a random stranger, *giving to charity* is the share of the endowment given the charity, *giving to government* is the share of the endowment given to the government program, *trust* indicates if the respondent invested in the partner, *ultimatum* indicates whether the proposer's offer was rejected, *inequality aversion* indicates if the respondent preferred the egalitarian allocation, *Meritocratic preferences* is the share of the endowment given to the subject who exerted effort, *impatience* is the premium that must be offered to the respondent to delay the payment for a year, *risk aversion* is the premium that must be offered to accept the risk, *tax morale* indicates in a 1-3 scale if evading taxes is justifiable with 3 being never justifiable, *Workers' evasion* indicates in a 1-10 scale the perceived share of wage earners who evade taxes (10 = 90-100%), *Firms' evasion* indicates in a 1-10 scale the perceived share of value added tax that is evaded by firms (10 = 90-100%), *trust in others* indicates in a 1-2 scale if most people can be trusted, *trust in government* indicates in a 1-5 scale whether one can trust the government in acting properly, *government efficiency* indicates in a 1-4 scale the perceived efficiency of the government, *preferences for redistribution* indicates in a 1-4 scale whether the government should take steps to reduce the income gap between rich and poor, *Left-Right spectrum* indicates in a 0-10 scale the respondents' placement in the left-right spectrum (10 = right), *perceived progressiveness* indicates the respondents' perceived difference in tax rate paid between the upper 20% and the bottom 20%, *desired progressiveness* indicates in a 1-4 scale whether the tax system should be more progressive, *perceived inequality* indicates in a 1-3 scale whether inequality is too high, *role of luck* indicates in a 0-2 scale if luck is important to determine incomes.

Table 3: Measurement Error Correction: Re-Scaling Approach

	Corr. Coef (1)	Scaling factor (2)	Source (3)	Rescaled Bounds (4)
Evasion aversion	0.008	0.09-0.69	Combined	[0.011,0.087]
Public goods	0.003	0.12-0.28	Carlsson et al. (2014)	[0.011,0.025]
Dishonesty	-0.000	0.09-0.69	Combined	[-0.003,-0.000]
Giving to strangers	0.040	0.09-0.48	Brosig et al. (2007)	[0.083,0.443]
Giving to charity	0.046	0.09-0.69	Combined	[0.067,0.514]
Giving to government	0.048	0.09-0.69	Combined	[0.070,0.538]
Trust	-0.008	0.354-0.69	Chuang and Schechter (2015) & Lönnqvist et al. (2015)	[-0.024,-0.012]
Ultimatum	0.026	0.09-0.69	Combined	[0.037,0.284]
Inequality aversion	0.007	0.09-0.69	Combined	[0.011,0.083]
Meritocratic preferences	-0.001	0.09-0.69	Combined	[-0.013,-0.002]
Impatience	0.021	0.4-0.67	Meier and Sprenger (2015) & Dean and Sautmann (2021)	[0.031,0.052]
Risk aversion	0.013	0.20-0.38	Levin et al. (2007)	[0.033,0.063]

Notes: Each row corresponds to one of the 12 lab-measures used in the paper. Column (1) corresponds to the correlation coefficient between each of the lab measures and the measure of tax evasion (reproduced from column (1) of Table 2). Column (2) presents the range of the scaling factor used in other papers. Column (3) identifies the data source for the scaling factor. "Combined" implies that lower and upper scaling factors are not available for these measures, so we take the highest and lowest values across all measures from Brosig et al. (2007) and Lönnqvist et al. (2015). Column (4) presents the rescaled bounds for the correlation coefficients.

Table 4: Measurement Error Correction: Obviously Related Instrumental Variables

			IV Strategy	
Endogenous variable (1)	Instrument(s) (2)	OLS Coeff. (3)	2SLS Coeff. (4)	F-stat (5)
Evasion aversion	Tax morale	0.008	-0.017	989.284
		[-0.017,0.033]	[-0.085,0.050]	
Giving to government	Trust in government & Government efficiency	0.048	0.105	650.189
		[0.023,0.074]	[0.045,0.165]	
Trust	Trust in others	-0.008	0.164	131.902
		[-0.033,0.017]	[-0.011,0.339]	

Notes: Estimates based on the Obviously Related Instrumental Variables approach for correcting measurement error. Each row corresponds to a different regression with a single right-hand-side variable, which is normalized to have a mean of 0 and a standard deviation of 1. The dependent variable is always an indicator variable for whether the taxpayer under-reported wages in 2016. *Evasion aversion* is the share of the endowment given to the partner who thinks evading taxes is never justifiable, *tax morale* indicates in a 1-3 scale if evading taxes is justifiable with 3 being never justifiable, *giving to government* is the share of the endowment given to the government program, *trust in government* indicates in a 1-5 scale whether one can trust the government in acting properly, *government efficiency* indicates in a 1-4 scale the perceived efficiency of the government, *trust* indicates if the respondent invested in the partner, and *trust in others* indicates in a 1-2 scale if most people can be trusted. Columns (1) and (2) presents the endogenous variable and the instrument(s) used in each case. Column (3) shows the OLS coefficients. Column (4) shows the 2SLS coefficients. 95% confidence intervals reported in brackets and based on robust standard errors. Column (5) present the the F-test of weak instruments for each 2SLS regression.

Table 5: Predicting Tax Evasion Choices with Multivariate Probit Regression

		(1)	(2)	(3)	(4)	(5)	(6)
Lab-Based	Evasion aversion	-0.002		0.002			
		(0.021)		(0.023)			
	Public goods	-0.007		-0.012			
		(0.022)		(0.022)			
	Dishonesty	0.000		0.000			
		(0.020)		(0.020)			
	Giving to strangers	0.041*		0.031			
		(0.021)		(0.022)			
	Giving to charity	0.067***		0.059***			
		(0.021)		(0.022)			
	Giving to government	0.072***		0.037			
		(0.020)		(0.023)			
	Trust	-0.025		-0.021			
		(0.021)		(0.021)			
	Ultimatum	0.026		0.021			
		(0.020)		(0.020)			
	Inequality aversion	0.004		0.009			
		(0.021)		(0.021)			
	Meritocratic preferences	0.013		0.021			
		(0.021)		(0.022)			
	Impatience	0.031		0.028			
		(0.021)		(0.021)			
	Risk aversion	0.011		0.015			
		(0.021)		(0.021)			
Survey-Based	Tax morale		-0.012	-0.021			
			(0.021)	(0.022)			
	Workers' evasion		0.020	0.021			
			(0.026)	(0.026)			
	Firms' evasion		0.092***	0.089***			
			(0.025)	(0.025)			
	Trust in others		0.020	0.026			
			(0.021)	(0.021)			
	Trust in government		-0.010	-0.011			
			(0.027)	(0.027)			
	Government efficiency		0.033	0.022			
			(0.026)	(0.027)			
	Preferences for redistribution		-0.003	-0.008			
			(0.025)	(0.025)			
	Left-right spectrum		-0.051**	-0.039			
			(0.026)	(0.026)			
	Perceived progressiveness		-0.004	0.002			
			(0.021)	(0.021)			
	Desired progressiveness		0.056**	0.053**			
			(0.026)	(0.026)			
	Perceived inequality		0.014	0.011			
			(0.023)	(0.023)			
	Role of luck		0.033	0.029			
			(0.023)	(0.023)			
	Past evasion				✓		
	Economic incentives					✓	
	Peer evasion						✓
	AUC	0.564	0.587	0.602	0.575	0.615	0.872
	Pseudo R^2	0.008	0.014	0.018	0.026	0.038	0.271
	Observations	6,078	6,078	6,078	6,078	6,016	6,078

Notes: Each column corresponds to a different Probit regression. Raw coefficients with robust standard errors in parentheses. The dependent variable indicates whether the taxpayer under-reported wages in 2016. All the survey-based and lab-based were normalized to have mean 0 and standard deviation 1 (for their definitions, see the notes to Table 2). *Past evasion* corresponds to a set of three variables on the individual's own tax evasion in 2015. *Economic incentives* correspond to a set of 7 dummies indicating the marginal tax rate the individual faces according to the level of the third-party reported salary, and a set of 9 decile dummies of the level of third-party reported salary. *Peer evasion* includes a set of three variables on the tax evasion of coworkers in 2016.