

# 40 Years of Tax Evasion Games: A Meta-Analysis

TARC DISCUSSION PAPER 031 - 20

JAMES ALM, TARC, UNIVERSITY OF EXETER  
ANTOINE MALEZIEUX, TARC, UNIVERSITY OF EXETER

# 40 Years of Tax Evasion Games: A Meta-Analysis <sup>\*</sup>

James Alm<sup>†</sup>     Antoine Malézieux<sup>‡</sup>

September 3, 2020

## Abstract

We collect individual participant data from 70 papers that use laboratory experiments to examine individual tax evasion behavior (or “Tax Evasion Games”), in order to use meta-analysis to estimate the impacts of different public policy, experimental design and individual level variables on tax evasion choices. Our results show that standard enforcement variables like audits (including audit rules) and fines perform differently on the extensive and intensive margins. We find that other fiscal variables like a flat tax system, tax rates, and tax amnesties have unambiguous negative impacts on tax compliance, and that specific features of the experimental setting, such as how subjects are directed to report income, or whether taxes are redistributed to the participants or to a real life public good, have significant impacts on tax compliance. Our results also indicate that the demographic characteristics of the subjects (e.g., gender, experimental income, occupation, risk attitude) affect compliance.

**Keywords:** tax evasion, tax compliance, meta-analysis.

**JEL classification:** C9, H0, H3.

---

<sup>\*</sup>We thank John D’Attoma, Duccio Gamannossi Degl’innocenti, Miguel Fonseca, Xavier Gassmann, Nicolas Jacquemet, Matthias Kasper, Stéphane Luchini, Daniele Nosenzo, Salmai Qari, Alessandro Santoro, the participants to the MUEES, TARC, Augustin Cournot, ADRES, Maastricht, Budapest, ESA, Lille seminars and all the researchers who contributed to this meta-analysis. We also thank three anonymous reviewers along with Ragan Petrie and Marie-Claire Villeval. The dataset generated in this meta-analysis is not publicly available as data are property from their respective authors. The dataset can be obtained on request, after express authorization from the authors who did not make their data available online. The authors declare no competing interests.

<sup>†</sup>Department of Economics, Tulane University, 208 Tilton Hall, New Orleans, LA 70118, USA (phone +1 504 862 8344; email [jalm@tulane.edu](mailto:jalm@tulane.edu)).

<sup>‡</sup>CEREN, EA 7477, Burgundy School of Business, Université Bourgogne Franche-Comté, 29 Rue Sabin, 21000 Dijon, France (email [antoine.malezieux@bsb-education.com](mailto:antoine.malezieux@bsb-education.com)).

# 1 Introduction

The year 2018 marked the 40th anniversary of the first article that used laboratory experiments to examine individual tax evasion behavior – what we term “Tax Evasion Games” (or TEGs). In this seminal contribution, Friedland, Maital, and Rutenberg (1978) asked 15 Israeli psychology undergrads to take part in the following experiment:

*“This research takes the form of an economic game. In general, each one of you will receive salary slips. You will be asked to report your income, and pay income tax on the income you reported. From time to time, audits will be conducted according to a random sample, and fines imposed on tax evaded. At the end of each “round” of 10 months, each person’s net income will be added up (gross income less income tax less fines). The objective of each person in the game is to accumulate the maximum amount of net income.”*

Since 1978, this simple game has been the subject of well over 130 publications (Torgler, 2016), with more papers added on a regular basis. This extensive use of TEGs has been driven by three main factors. First, there are compelling reasons to find ways to fight tax evasion, and laboratory experiments offer one way to examine these policy tools. Second, laboratory experiments generate observable and reliable data about individual tax evasion choices, information that is by nature very difficult to measure in the field. Third, laboratory experiments have a high degree of “internal validity”, allowing researchers to test hypotheses that isolate the cause-and-effect of the policy intervention.

However, despite the wide use of TEGs, it is disappointing – and surprising – that the impacts of many variables examined in these TEGs remain unclear (Muehlbacher and Kirchler, 2016; Malézieux, 2018). It may seem obvious that increasing the audit probability and/or the penalty rate will increase compliance, but even so these impacts may be different on the intensive margin (i.e., the compliance from evaders) versus on the extensive margin (i.e., the probability of full compliance). The lessons are even more uncertain when we look at other policy levers like the tax rate, the audit rule, the type of tax system, or the use of tax amnesties. There is also uncertainty about the ways in which subject demographic characteristics affect compliance behavior, and the impacts of such experimental design features as the nature of experimental instructions, the presence or absence of an earning task, the ways in which subjects are directed to report income, or the redistribution of taxes to the participants or to a real life public good, have a significant impacts on tax compliance.

As a result, we believe that it is worthwhile to determine what we have learned from these TEGs, as well as what we have still to learn from them. These are our goals here: To examine the impacts on tax evasion behavior in TEGs of traditional public policy variables (e.g., the audit rate, the fine rate, the tax rate, the audit rule, a tax amnesty, the tax system), experimental design variables (e.g., the framing of the instructions, the use of an earning task, the redistribution of

taxes), and individual-level variables (e.g., age, gender, occupation, risk attitude), and to identify fruitful areas for future TEG research.

We do this by collecting 70 TEG datasets and then using meta-analysis to estimate these impacts. As defined by Glass (1976), a meta-analysis is “...*the statistical analysis of a large collection of [...] results from individual studies for the purpose of integrating the findings*” (p.3). The term “meta-analysis” refers to “the analysis of analyses”. This method has been extensively used in clinical research, and it has attracted increasing attention in experimental social sciences (Davis, Mengersen, Bennett, and Mazerolle, 2014). The aim of a meta-analysis is to determine, if possible, the definitive impact of a variable, to guide the use of replication analyses, and to inform public policy (Maki, Cohen, and Vandenberg, 2018).

There are two types of meta-analysis, one based on aggregated data and another based on individual participant data (Riley, Lambert, Staessen, Wang, Gueyffier, Thijs, and Boutitie, 2008). In an important earlier paper on tax evasion behavior in TEGs, Blackwell (2010) ran a meta-analysis on aggregated TEG data, studying the compliance impact of the audit probability, the fine rate, the tax rate, and the multiplier on the public good. However, Blackwell (2010) was only able to use 26 articles in his work, given the time at which he conducted the meta-analysis, and he was also unable to differentiate between the intensive and the extensive margins. The considerable expansion in TEGs over the last decade allows us to increase considerably the number of articles included in our meta-analysis, as well as to increase the number of variables that we can examine (including intensive and extensive margins). We are also able to utilize individual participant data, a method considered the “gold standard” of meta-analysis, with many advantages over meta-analysis of aggregated data (Riley, Lambert, Staessen, Wang, Gueyffier, Thijs, and Boutitie, 2008).

We collect an enormous dataset of 256,801 observations on 16,467 subjects, more than any previous study on tax evasion behavior. Our results for policy levers show that audits and fines interact differently on the extensive and intensive margins. Those variables have a positive impact on the extensive margin, and their interaction is positive; that is, the higher the audit probability, the greater is the impact of the fine rate on the decision to comply, and vice versa. However, their interaction is negative on the intensive margin, so that the higher the audit probability, the more negative is the impact of the fine rate for tax compliance coming from evaders, and vice versa. In addition, we find that a flat tax system, the tax rate, and a tax amnesty have an unambiguously negative impacts on tax compliance. We also find that two endogenous audit rules achieve higher compliance than a random audit rule, while other endogenous audit rules have no significant impact, suggesting a mixed performance of endogenous audits. As a comparison, the original Allingham and Sandmo (1972) model of tax compliance predicts that increases in the audit probability and the fine size both increase tax compliance. The present meta-analysis indeed finds that audits and fines both have a positive impact on the extensive margin. However, we also find that on the intensive margin audits and fines work against each other. Also, contrary to the predictions of Yitzhaki (1974), there is a negative impact of raising tax rates, at least on

the extensive margin of compliance. As for design features, we find no impacts of framing the experiment with non-neutral terms, directing subjects to report income or of making participants earn an income rather than giving them income. We also find that redistributing the taxes that are collected to the participants or to an investment in a real life public good each has a positive impact on tax compliance. We also show that using a research fund as a real life public good is a representative type of public good to implement in a TEG. Finally, our estimation results indicate the importance of some individual-level characteristics. The typical tax evader in the lab is a male who is risk averse. On average, the higher his income in Euros, the higher is his probability of evading. Being a student increases his share of evasion. We do not find any impact from the subject's age.

The outline of the paper is as follows. Section 2 provides a literature review of all the variables under study here. Section 3 reports on our data and our methods for conducting a meta-analysis, and Section 4 presents the results of our meta-analysis. We conclude in Section 5.

## 2 Literature review

We review here the literature on the most prominent variables and their expected impact on tax compliance in the laboratory. These variables are split into three categories: the main public policy variables, the experimental design variables, and the individual-level variables.

### 2.1 Public policy variables

It is useful at the start to examine what theory tells us about the impact of public policy variables on individual tax evasion. The basic theoretical model used in nearly all research on tax compliance starts with the work of Becker (1968) and his economics-of-crime model, as first applied to tax compliance by Allingham and Sandmo (1972) and Srinivasan (1973). Here a rational individual is viewed as maximizing the expected utility of the tax evasion gamble (or lottery), weighing the expected benefits of successful cheating against the risky prospect of detection and punishment. The standard conclusion from this approach is that an individual pays taxes because he or she is afraid of getting caught and penalized if he or she does not report all income. This approach therefore gives the plausible and productive result that compliance depends upon audit and fine rates. Indeed, the central point of this approach is that an individual pays taxes because – and only because – of this fear of detection and punishment.

To illustrate this expected utility approach more precisely, consider a simple version of the standard model. An individual is assumed to receive a fixed amount of income  $I$ , and must choose how much of this income to report to the tax authorities and how much to underreport. The individual pays taxes at rate  $t$  on every dollar  $R$  of income that is reported, while no taxes are paid on underreported income. However, the individual may be audited with a fixed probability  $p$ ; if audited, then all underreported income is discovered, and the individual must pay a penalty

at rate  $f$  on each dollar that he was supposed to pay in taxes but did not pay. The individual's income  $I_C$  if caught underreporting equals

$$I_C = I - tR - f[t(I - R)] \quad (1)$$

or income less taxes paid on reported income less penalties on unreported taxes. If underreporting is not caught, income  $I_N$  is

$$I_N = I - tR \quad (2)$$

or income less taxes paid on reported income. The individual is assumed to choose declared income to maximize expected utility, defined as

$$EU(I) = pU(I_C) + (1 - p)U(I_N) \quad (3)$$

where  $E$  is the expectation operator and utility  $U(I)$  is a function only of income. This optimization generates a standard first-order condition for an interior solution; given concavity of the utility function, the second-order condition is satisfied. Comparative statics results are easily derived. For example, it is straightforward to show that an increase in either the probability of detection  $p$  or the penalty rate  $f$  unambiguously increases reported income.

This economics-of-crime approach therefore gives the sensible result that compliance depends upon enforcement. Indeed, it is essential to recognize that this approach concludes that an individual pays taxes because – and only because – of the economic consequences of detection and punishment. This is an important insight, with the obvious implication that the government can encourage greater tax compliance by increasing audit and penalty rates. Perhaps surprisingly, this approach also predicts that a higher tax rate has an ambiguous effect on compliance, as discussed later. Indeed, a common theoretical result is that a higher tax rate actually increases compliance, at least under plausible assumptions about how an individual is penalized on detected evasion and how the individual evaluates risk (Yitzhaki, 1974).

This basic model has been considerably expanded over the years to incorporate a range of additional factors that seem relevant to the tax evasion decision, including such factors as endogenous audit rules, alternative tax systems, and a tax amnesty, among many other considerations. The basic model has also been expanded to incorporate behavioral economics considerations. See Alm (2019) and Slemrod (2019) for comprehensive surveys of the tax compliance literature.

This model, along with its many extensions, forms the basis for most TEGs, often with clear predictions about the impacts of various public policy variables. Consider these variables and their predicted effects.

This model assumes that audits are always random (i.e., that each taxpayer has the same chances of getting audited regardless of what is reported). However, the laboratory allows researchers to test alternative audit rules. The type of audit can either be random or strategic. A

strategic (or “endogenous”) audit rule means that the tax administration can utilize information conveyed by the taxpayer to select, or to target, strategically a sub-sample of taxpayers. For example, an endogenous audit rule may be related to the relative behavior of a taxpayer within a group or to the compliance history of the taxpayer. The literature typically demonstrates that endogenous audit selection generates more compliance than random audit selection (Malézieux, 2018).

In the laboratory, the audit probability represents the chances that participants’ declarations will be checked, discovered, and fined. As the probability of audit raises, the basic model (and nearly all of its many extensions) predicts that taxpayers should be less willing to evade. Indeed, the audit rate has arguably the most consistent and positive impact on compliance in TEGs. As indicated by Malézieux (2018), about 80% of the TEGs find that increasing the audit probability, all else equal, increases tax compliance. The previous meta-analysis from Blackwell (2010) also finds this result.

In real life, taxpayers typically do not know the objective audit probability that they face, in large part because tax administrations rarely publicize this information. There is some theoretical work on the effects of this ambiguity (Alm, 1988, 2014; Snow and Warren, 2005), but without much resolution. The few laboratory experiments on this issue (Friedland, 1982; Alm, Jackson, and McKee, 1992; Choo, Fonseca, and Myles, 2015) also provide little resolution.

The fine rate on detected evasion is another important policy variable. As with the audit rate, the basic model and its extensions predict that an increase in the fine rate will increase taxpayer compliance. Malézieux (2018) finds that 70% of the TEGs generates this effect. Blackwell (2010) also estimates that the fine rate has a positive impact on tax compliance.

A tax amnesty is a program in which a participant who has previously evaded taxes may pay these unpaid taxes without being subject to the additional fines that detected evasion typically brings. Alm, McKee, and Beck (1990) implement an amnesty in which participants may pay some or all of their undeclared taxes from previous rounds, without any fine imposed. Rechberger, Hartner, Kirchler, and Hämmerle (2010) run a similar experiment, in which evaders may engage in an amnesty program that will reduce their tax rate on previously undeclared income. Both papers show that a tax amnesty often has positive impacts on post-amnesty compliance, depending on the specific amnesty features. Torgler and Schaltegger (2005) implement exactly the same amnesty system as in Alm, McKee, and Beck (1990). Like Alm, McKee, and Beck (1990), they also find that an amnesty increases tax compliance, but in contrast they find that an amnesty may also reduce post-amnesty compliance, depending on its specific features. These studies also examine whether the expectation of an amnesty affects compliance. Alm, McKee, and Beck (1990) find that expecting an amnesty reduces compliance, and Torgler and Schaltegger (2005) find that the expectation decreases the positive impact of an amnesty.

The tax rate determines the amount of subjects’ earnings that the subject is expected to pay. As noted earlier, its impact on compliance is theoretically ambiguous. As first shown by

Allingham and Sandmo (1972), the tax rate has two effects on compliance. An increase in the tax rate reduces compliance via a substitution effect (e.g., an increase in the tax rate reduces the return to compliance); however, an increase in the tax rate also reduces income, with effects on compliance that depend upon the individual’s attitude toward risk. In fact, when an individual exhibits decreasing absolute risk aversion, an increase in the tax rate is predicted to increase compliance because the higher tax rate reduces income, making the taxpayer more risk averse, and the taxpayer responds by reporting more income. The net impact of the tax rate on compliance is therefore ambiguous, dependent upon the strength of the substitution and the income effects. However, this theoretical ambiguity also depends upon the precise way in which the penalty is imposed. As shown by Yitzhaki (1974), when the imposed penalty is set on evaded taxes (rather than evaded income), an increase in tax rate induces more tax compliance because the substitution effect disappears and only the income effect remains. It is also possible to construct a theoretical model in which an increase in the tax rate reduces compliance, as shown by Bernasconi, Corazzini, and Seri (2014) and, more generally, by Hashimzade, Myles, and Tran-Nam (2013). Roughly 60% of the papers cited in Malézieux (2018) show that there is a negative link between tax rate and compliance. Even so, Blackwell (2010) finds that tax rate had an insignificant impact on compliance. Andreoni, Erard, and Feinstein (1998) concluded that, given its importance, the effect of the tax rate “...surely deserves further investigation” (p. 839).

A flat rate tax system is a system where the participants are all taxed at the same constant tax rate  $t$ , whatever their income. When the tax system is progressive, participants earning different income face different tax rates. There is little work on this issue. Heinemann and Kocher (2013) show that participants evaded more under a flat rate tax than under a progressive tax system. One explanation for this result is that a progressive tax system increases the tax rate for most taxpayers, which should in turn increase tax compliance (Yitzhaki, 1974). Another explanation relates to any perceived unfairness of a flat tax system, as found in an experimental paper by Spicer and Becker (1980).

## 2.2 Experimental design variables

There are various design features that may affect the results of the TEGs.

The framing of the experiment has been recognized as one of the key design elements of a TEG, to ensure some validity to the experiment (Torgler, 2002; Muehlbacher and Kirchler, 2016). A “loaded” frame uses non-neutral terms to place the experiment squarely in the tax context, such as tax compliance-related terms like the tax rate, the audit rate, the penalty rate, reported income, and so on. The alternative is a “neutral” frame with more value-free terms (e.g., payment, check, shortfall, disclosed money). Numerous experiments have shown that framing increases tax compliance (Baldry, 1986; Webley and Halstead, 1986; Wartick, Madeo, and Vines, 1999; Mittone, 2006; Trivedi and Chung, 2006; Choo, Fonseca, and Myles, 2015). However, this effect is not automatic (Alm, McClelland, and Schulze, 1992), and it could depend on the interaction of

framing with other variables (Durham, Manly, and Ritsema, 2014).

The way by which to request participants are asked to pay taxes also matters in a TEG, the request can be more or less directive. Cadsby, Maynes, and Trivedi (2006) show that, when participants are presented a classical way to request a declaration, subtly inviting them to declare the amount they want between 0 to 100% of their income, compliance rates are around 50%. In another condition, when a directive injunction to declare taxes is used, suggesting “obedience to authority”, they find that compliance rates increase drastically to almost 100%.

Mirroring real life settings, most experiments ask participants to declare an income. However, several papers request participants to directly declare the taxes that they want to pay (Kastlunger, Kirchler, Mittone, and Pitters, 2009; Kastlunger, Dressler, Kirchler, Mittone, and Voracek, 2010; Kastlunger, Muehlbacher, Kirchler, and Mittone, 2011; Hartl, Hofmann, Gangl, Hartner-Tiefenthaler, and Kirchler, 2015; Kogler, Mittone, and Kirchler, 2016; Mittone, Panebianco, and Santoro, 2017; Lambertson, De Neve, and Norton, 2018). To our knowledge, no article has directly compared both ways of running a TEG.

In an experiment, income can be earned by the participant or endowed by the experimenter to the participant. Investing effort to earn an income is often thought to make participants more willing to evade (via a sunk cost effect, a feeling of ownership or entitlement). However, any effort invested could also increase risk averse decisions, making participants less willing to evade via a reverse sunk cost effect (Durham, Manly, and Ritsema, 2014). Despite its recognized importance (Torgler, 2002; Muehlbacher and Kirchler, 2016), there is no clear effect either of the origin of income on tax compliance decisions or of the level of difficulty in earning an income (Malézieux, 2018).

The use of any taxes that are collected from participants is also a relevant design feature. For example, taxes can be redistributed directly back to the participants, with a positive social multiplier. Such redistribution has a strong positive impact on laboratory tax compliance (Becker, Büchner, and Slesking, 1987; Alm, Jackson, and McKee, 1992; Alm, McClelland, and Schulze, 1992; Alm, Sanchez, and De Juan, 1995; Bosco and Mittone, 1997; Alm, McClelland, and Schulze, 1999; Park and Hyun, 2003; Torgler, 2003; Gërxhani and Schram, 2006; Blackwell, 2010). Alternatively, taxes can also fund a real life public good, such as a donation to an organization, a student scholarship, or a research fund. The impact of a real life public good on compliance has been found to vary across its implementation, and no definitive answer has yet emerged (Mittone, 2006; Masclet, Montmarquette, and Viennot-Briot, 2013; Malézieux, 2018). However, when Doerrenberg (2015) compared investing taxes in real life public goods or redistributing taxes to participants, he found no difference in the results.

### **2.3 Individual-level variables**

Numerous experiments have found that older subjects tend to be more tax compliant (Coricelli, Joffily, Montmarquette, and Villeval, 2010; Eisenhauer, Geide-Stevenson, and Ferro, 2011; Dulleck,

Fooken, Newton, Ristl, Schaffner, and Torgler, 2016; Mittone, Panebianco, and Santoro, 2017; Muehlbacher, Hartl, and Kirchler, 2017; Vossler and McKee, 2017). However, this result is not universal (Boylan, 2010; Rechberger, Hartner, Kirchler, and Hämmerle, 2010; Lefebvre, Pestieau, Riedl, and Villeval, 2015). Abeler, Nosenzo, and Raymond (2019) do not find any impact of age on cheating in a meta-analysis of lying games whereas Engel (2011) shows that age is associated with an increase with generosity in dictator games.

Several studies argue that men behave more dishonestly than women with respect to taxes (Castro and Rizzo, 2014; Coricelli, Rusconi, and Villeval, 2014; Tan and Yim, 2014; Dulleck, Fooken, Newton, Ristl, Schaffner, and Torgler, 2016; Mittone, Panebianco, and Santoro, 2017; Abraham, Lorek, Richter, and Wrede, 2017; Vossler and McKee, 2017; D’Attoma, Volintiru, and Malézieux, 2020). This result has also been found in meta-analyses (Engel, 2011; Abeler, Nosenzo, and Raymond, 2019). The underlying factors could be that women are more risk-averse (Borghans, Heckman, Golsteyn, and Meijers, 2009), that they exhibit more pro-social behavior (Grosch and Rau, 2017), or that they are compensating for worse performance in tasks preceding the tax evasion experiments (Jacobsen, Fosgaard, and Pascual-Ezama, 2017).

TEGs are most often run with student subjects, which is often cited as a main concern about the external validity of those experiments (Levitt and List, 2007; Alm, Bloomquist, and McKee, 2015). Numerous experiments have been run exclusively with student samples of subjects, even though an increasing number are now run with non-students. Papers that have compared those different population have found varied results on student versus non-student compliance (Gërkhani and Schram, 2006; Choo, Fonseca, and Myles, 2015; Alm, Bloomquist, and McKee, 2015). Student have been found to cheat more than non-students in a meta-analysis on lying games (Abeler, Nosenzo, and Raymond, 2019), and students have also been found to be less generous than non-students in dictator games (Engel, 2011).

Other individual-level variables include subject income. It is often found in real world data that there is a positive correlation between taxpayer’s income and tax evasion because people may be more willing to engage in risky activities as they get richer (Johns and Slemrod, 2010; Alstadsæter, Johannesen, and Zucman, 2019). However, the impact of income on tax evasion in the laboratory is not as certain, ranging from no effect (Gërkhani and Schram, 2006; Hsu, 2008; Jacquemet, Luchini, Malézieux, and Shogren, 2020), to a positive effect (Heinemann and Kocher, 2013; Coricelli, Rusconi, and Villeval, 2014; Lefebvre, Pestieau, Riedl, and Villeval, 2015), and even to a negative effect (Bruttel and Friehe, 2014; Choo, Fonseca, and Myles, 2015). The way in which the tax reporting is framed (e.g., gain versus loss) may also play a role here. Note that Abeler, Nosenzo, and Raymond (2019) find that an increase in the stakes of lying games does not change the average amount of lying.

Individual risk attitudes are also expected to affect tax evasion choices. Complying gives a certain outcome whereas evading depends on the occurrence of an audit and a fine, so there should be a positive relationship between a taxpayer’s level of risk aversion and tax compliance. In the

laboratory, the standard way of capturing risk attitude is to ask participants to choose between a series of safe and risky lotteries (Holt and Laury, 2002). However, there is no uniform result on risk attitudes in TEGs, with studies finding a positive correlation (Heinemann and Kocher, 2013; Guerra and Harrington, 2018; Vossler and Gilpatric, 2018), a negative correlation (Bühren and Kundt, 2014; Choo, Fonseca, and Myles, 2015), and no correlation (Dulleck, Fooker, Newton, Ristl, Schaffner, and Torgler, 2016; Abraham, Lorek, Richter, and Wrede, 2017). Other measures, such as one-item self-assessments, other type of lotteries, personality questionnaires, have been used with no definitive results.

The next section discusses our data and our methods for conducting a meta-analysis on the impact of those variables on laboratory tax compliance.

### 3 Data and methodology

Our data collection follows procedures recommended by the PRISMA-IPD, as developed by Stewart, Clarke, Rovers, Riley, Simmonds, Stewart, and Tierney (2015).

The first step is to identify the articles to include. We started with the TEGs cited by Torgler (2016) in a previous survey of tax evasion studies identified through database searching on Econlit, Web of Science, and SCOPUS. We then added other sources by reading different literature reviews, consulting their bibliographies, using different resumes of researchers, and directly contacting researchers. The resulting article collection was first used in the literature review published in Malézieux (2018).

The second step is the definition of an eligibility criterion and a screening of the studies. We excluded all the articles that did not implement a TEG (i.e., all the articles that did not focus on the task of declaring income for the purpose of taxation). We added two Public Good Games to this list because their main tasks were very close to the one implemented in a TEG (Silverman, Slemrod, and Uler, 2014; Špalek and Špačková, 2016). We ended up with a list of 137 possible articles, from 1978 till 2018.<sup>1</sup> All except one use monetary incentives (Peliova, 2015).

Collection – the third step – started in December 2017, and ended in February 2019. After the article collection, we retrieved the email address of each corresponding author, and we sent emails to the current known author, asking about the status of the dataset (available or not) and the possibility of sharing the dataset with us. If there was no initial response, we sent reminders (on average 4 reminders). Very few authors chose to not answer (or to stop answering) our requests ( $N = 8$ ).<sup>2</sup> Out of the 137 articles, the data from 70 of these articles were successfully retrieved, while 59 of them were considered as lost, with most of these 59 lost articles representing earlier and so older papers. Authors from 7 articles refused to share their data, mainly because the data were not published yet.<sup>3</sup> One dataset was judged as unusable because the data were not in a

---

<sup>1</sup>The evolution of the number of TEG published per year is shown in Figure 2 in the Appendix.

<sup>2</sup>For copies of the email data requests and reminders, see Figures 4 and 5 in the Appendix.

<sup>3</sup>The estimated numbers of participants lost due to those collection problems is at most equal to 7835, which

usable state and the instructions for the experiment were lost.<sup>4</sup>

Another important feature of a meta-analysis is the way to code the data obtained, which is probably the most time-consuming step in a meta-analysis on individual participants. We describe below the twenty variables under study. When the information was missing or undetermined, it was systematically coded as missing.

**Audit rule:** The type of audit is either *random* – each participant has the same probability of being audited – or *strategic*. The alternative strategic audit types are the following: *Risk-based audit rule* (an individual audit probability is computed for each taxpayer, according to his estimated chances of evasion, e.g., Vossler and Gilpatric, 2018); *Coordination audit rule* (the audit works as a coordination game, e.g., Tan and Yim, 2014; Dai, 2016; Etchart-Vincent and Taugourdeau, 2018); *Cutoff audit rule* (participants’ declarations below a certain threshold are audited, e.g., Coricelli, Joffily, Montmarquette, and Villeval, 2010; Coricelli, Rusconi, and Villeval, 2014); and *Invest audit rule* (participants can invest a part of their tax revenue on auditing, e.g., Hsu, 2013). This variable is coded as a dummy variable equal to 1 if the audit is random and 0 otherwise. The effects of strategic audit rules are examined separately, as discussed later.

**Audit probability:** This variable represents the chances that participants’ declarations will be checked and fined. It is coded as a percentage ranging between 0 to 1.

**Fine size:** This variable represents the amount of penalty applied to the participant’s unreported taxes if caught evading. This number is computed as a multiple of the amount of taxes evaded, ranging from 0 to 10 and excluding the initial reimbursement of the taxes evaded.

**Tax amnesty:** There is a *tax amnesty* when participants who have previously evaded taxes may pay these unpaid taxes without being subject to additional fines. It is coded as a dummy variable equal to 1 for the articles in which a tax amnesty is present (including when the fine rate is equal to zero) and 0 otherwise.

**Tax rate:** The tax rate determines the amount of subject’s earnings that the subject is expected to pay. It is coded as a percentage ranging between 0 to 1.

**Tax system:** The tax system implemented in the laboratory can either be *flat* (proportional) or *progressive* tax rate system.<sup>5</sup> In a flat tax system, all the participants are taxed at the same constant tax rate. In a progressive tax system, participants with higher income face a higher unique tax rate that is applied to their income. A flat rate tax system is coded as a dummy variable equal to 1 and 0 otherwise.

**Loaded framing:** The instructions presented to the participants can either be *loaded* (or tax-related), or *neutral*. A loaded framing is coded as a dummy variable equal to 1 and 0 otherwise.

**Directive way:** The instructions presented to the participants can be more or less directive in

---

represents 102,191 individual subject decisions. However, these numbers are likely overestimates of the actual lost observations because many different articles have been published with a single dataset.

<sup>4</sup>The collection output is featured in Figure 3 in the Appendix.

<sup>5</sup>In principle, the tax system could of course be *regressive* as well. To our knowledge, there are no studies so far with a *regressive* tax system.

matter to the declaration request. An example of directive instructions could be as follows: “*You have to declare 100% of your income*”. An example of non-directive instructions could be: “*You may declare between 0 and 100% of your income*”. Directive instructions are coded as a dummy variable equal to 1 and 0 otherwise.

**Tax task:** Some experiments have asked participants to declare directly taxes rather than declaring an income. When the task is about declaring taxes directly, this variable is coded as a dummy variable equal to 1 and 0 otherwise.

**Earned income:** Income can either be *earned* by participants or *given* to participants. When income is earned, this variable is a dummy variable equal to 1 and 0 otherwise.

**Redistribution:** This dummy variable is coded as 1 when taxes collected are *redistributed* directly to the participants (whatever the size of the group multiplier) and 0 otherwise.

**Real life public good:** This variable is a dummy variable coded as 1 when taxes are invested in a real life *public good* (see below) and 0 otherwise. This variable is not nested or related in any way to the Redistribution variable.

**Real life public good type:** There are different types of real life public goods: *Research*, when the taxes that are collected from participants are kept by the experimenter (Coricelli, Joffily, Montmarquette, and Villeval, 2010; Coricelli, Rusconi, and Villeval, 2014; Adres, Vashdi, and Zelmanovitch, 2016); *Different*, when there are several organizations who receive the taxes (D’Attoma, 2018; D’Attoma, Volintiru, and Malézieux, 2020; Brockmann, Genschel, and Seelkopf, 2016); *Organization*, when the taxes are distributed to non-student organizations (Doerrenberg and Duncan, 2014b; Malézieux, 2016); *Student*, when the taxes are given to student organizations or for student purposes (Lamberton, De Neve, and Norton, 2018); or *Government*, when the taxes are donated to a governmental structure (Doerrenberg, 2015). These variables are all coded as dummy variables. Their separate effects are discussed later.

**Age:** The *age* variable represents the age of the participant at the time of the experiment. It is a positive integer ranging from 13 to 85.

**Male:** This dummy variable represents the *gender* of the participant, coded 1 for males and 0 for females.

**Student:** This variable measures the *subject pool*, student versus non-student subjects. A student subject is coded as a dummy variable equal to 1 and a non-student is coded as 0.

**Income:** This variable represents participants’ income level in the experiment. Most experiments use an Experimental Currency Unit (or ECU). Each income value has been converted into the currency used in the corresponding country, then exchanged for updated Euros from 2019. It is thus a positive integer ranging from 0 to 47.6 €.

**Income (PPP):** This variable also represents participants’ income level in the experiment. However, it takes into account Purchasing Power Parity (PPP). Each income value has been converted into the currency used in the corresponding country, exchanged for dollars eliminating the differences in price levels between countries, and finally exchanged for updated Euros from

2019. It is thus a positive integer ranging from 0 to 49.76 €.

**Risk attitude:** Following Holt and Laury (2002), this risk attitude variable represents the number of safe lotteries chosen. It ranges from 0 (0 safe choice made) to 10 (all safe lotteries selected). Another related variable that captures risk attitudes is **Risk averse (HL)**, where participants are considered risk averse if they have selected strictly more than 5 safe lotteries at the Holt and Laury (2002) task, in which they are coded as a dummy variable equal to 1 and 0 otherwise. The effects of different measures of risk attitudes are discussed later in our robustness tests.

## 4 Results

Summary statistics from the meta-analysis raw data are shown in Table 1, which indicates the name of the variable, the mean value, the standard deviation, the minimum value, the maximum, and the number of observations. As our results depend on the information collected in each original study, we observe first that the number of observations is not always the maximum for each variable, a problem addressed below.

### 4.1 Overall compliance

The rate of tax compliance is the common variable of interest in all of the studies aggregated here. It is defined as the ratio of declared income to true earned or received income, ranging from 0 to 1. Some studies allowed participants to declare more than their actual income (e.g., Gërxhani and Schram 2006), leading to compliance being higher than 1. In these cases, we set compliance equal to 1 for any declaration above 100% of one’s income.<sup>6</sup>

Table 1 reports summary statistics on compliance measures in the TEG meta-analysis. On average, compliance is equal to 65% (with a standard deviation of 41%). There is generally a tendency toward all-or-none behavior in TEGs (Alm, McClelland, and Schulze, 1992; Alm, Bloomquist, and McKee, 2015; D’Attoma, Volintiru, and Malézieux, 2020). The proportion of full compliance observations is equal to 45%, while 19% of the decisions are equal to full evasion. This bimodal distribution is confirmed by Figure 1, which clearly indicates that the two extremes are the two modes. The third mode – declarations corresponding to half of the income – is no higher than 5%.

### 4.2 Multivariate analysis

We turn now to parametric models to study the impacts of our different explanatory variables. It is standard to distinguish the extensive margin from the intensive margin of compliance. The first outcome is specified as a 0/1 variable, on which we estimate a Probit model. The second outcome

---

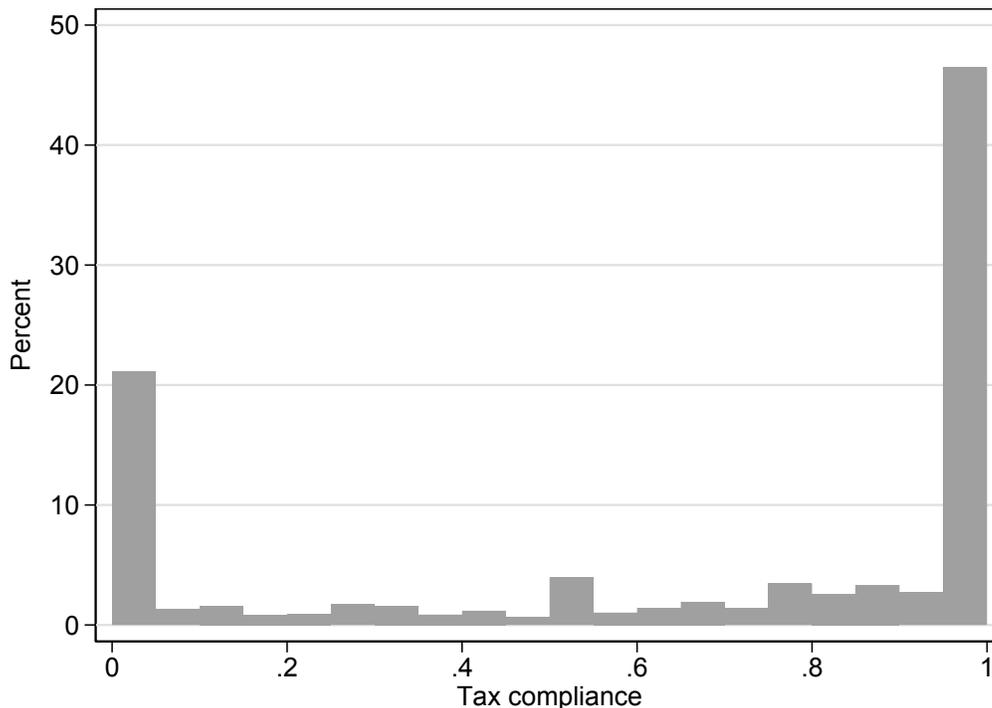
<sup>6</sup>Simply dropping those observations does not change the conclusions from the main analysis of this article.

Table 1: Summary statistics of compliance, public policy variables, experimental design variables, and individual-level variables

Variable	Value	SD	Min	Max	# Obs.	# Datasets
Number of observations	256,801	–	–	–	–	–
Number of subjects	16,467	–	–	–	–	–
Number of rounds	15.19	13.59	1	60	256,128	–
Countries	19	–	–	–	256,801	–
Compliance rate	.65	.41	–	–	219,545	–
Full compliance	.45	.49	–	–	256,801	–
Audit probability	.20	.15	0	.90	219,116	–
Fine size	2.01	1.47	0	10	217,799	–
Amnesty	.02	.15	–	–	217,799	3
Tax rate	.34	.11	0	.80	244,412	–
Flat tax	.96	.19	–	–	249,521	62
Audit: Random	.88	–	–	–	209,137	57
Audit: Risk-based	.07	–	–	–	17,096	1
Audit: Coordination	.03	–	–	–	7,140	3
Audit: Cutoff	.01	–	–	–	2,400	2
Audit: Invest	.007	–	–	–	1,853	1
Loaded framing	.89	.31	–	–	245,768	49
Directive way	.46	.49	–	–	159,870	12
Tax task	.33	.47	–	–	256,801	14
Earned income	.39	.48	–	–	253,601	27
Redistribution	.21	.41	–	–	256,801	18
Real life public good	.07	.26	–	–	250,442	16
RL public good: Research	.03	–	–	–	9,501	8
RL public good: Different	.02	–	–	–	5,779	3
RL public good: Organization	.007	–	–	–	1,847	4
RL public good: Student	.001	–	–	–	378	2
RL public good: Government	.001	–	–	–	279	1
Age	24	7.90	13	85	200,625	–
Male	.47	.49	–	–	209,204	–
Student	.91	.27	–	–	212,726	–
Risk: Holt & Laury	5.82	1.94	0	10	49,774	11
Risk averse (HL)	.58	.49	–	–	49,774	11
Income (€)	3.61	5.98	0	47.6	190,222	–
Income (PPP, €)	4.18	6.90	0	49.7	182,562	–

**Note.** Summary statistics on compliance, public policy, experimental design and individual-level variables. From left to right are the variable's name, mean or sum value, standard deviation, minimum, maximum and total number of observations for each variable. The last column represents the number of datasets in which the information concerning categorical variables is available.

Figure 1: Distribution of tax compliance rates



**Note.** The graph depicts the distribution of tax compliance decisions. Compliance is the ratio between income declared and total income.

is the ratio of the declared income to income earned or given, estimated using a linear model on evaders only (i.e., conditional upon some evasion).<sup>7</sup> The models are estimated with standard errors clustered at the individual level to capture possible intra-individual heterogeneity of compliance.<sup>8</sup> Our models are also estimated with several different types of fixed effects, accounting for any possible confounding factors related to the study (Study Fixed Effects), to the round (Round Fixed Effects), to the year (Year Fixed Effects), and to the country in which the experiment

<sup>7</sup>Those models are standard when it comes to analysing experimental data (e.g., Jacquemet, Luchini, Malézieux, and Shogren, 2019) or meta-analyses data (e.g., Abeler, Nosenzo, and Raymond, 2019). As discussed later, we have also estimated these specifications using a Logit model and a GLM model in Table 15 in the Appendix, and our results are not affected. For the sake of transparency, we have also reported in Table 16 in the Appendix, the average effect of the same variables using both Tobit and OLS models without distinguishing the extensive from the intensive margin of compliance.

<sup>8</sup>Clustering at the individual level is standard because we are mostly dealing with individual-level decision making. As discussed later, we have also estimated these specifications with clustering at the study level in order to capture possible intra-study heterogeneity of compliance; see Table 11 in the Appendix. Our results are unaffected. Another alternative would be to cluster at the session level. However, this information was not available in all the datasets of TEG that we gathered.

takes place (Country Fixed Effects).<sup>9</sup> Finally, because we do not have the maximum number of observations for each variable, pooling all the variables in only one regression would lead to a regression with an unacceptably small number of observations and to numerous omitted variables because of collinearity problems. As the aim of a meta-analysis is to investigate the effects from a significant number of observations coming from a maximum number of studies, we instead specify three models: one for the main public policy variables, one for the experimental design variables, and one for the individual-level variables.

#### 4.2.1 Public policy variables

We first investigate the multivariate impacts of our main public policy variables (audit, fine, and tax rates). We want to find the consistent impact, if any, of those public policy variables on tax compliance. The results are reported in Table 2, Column 1 for the extensive margin and Table 2, Column 4 for the intensive margin. We focus on each variable.

Having a random audit rule as compared to a strategic one has the expected negative sign on the extensive margin, but it is not significant ( $p = 0.211$ ).<sup>10</sup> On the intensive margin, the sign points in the opposite direction, while it is still not significant ( $p = 0.097$ ). This result is surprising because theory is clear that strategic audit rules are better able to deter tax evasion than random audit rules. Three problems might explain these results. First, our sample is highly unbalanced as about 88% of the audits here are random. Second, our sample is also highly heterogeneous because in the remaining 12% of the sample four different endogenous audit rules are considered, in which the experimental implementation of the audit rule is different for each article; that is, a coordination audit is not implemented using the same features in Tan and Yim (2014), in Dai (2016), or in Etchart-Vincent and Taugourdeau (2018). Third, out of the very few papers that implemented an endogenous audit, only two reported the objective audit probability that participants were facing, and both of these papers examined audits working as coordination games. This result therefore needs to be interpreted with caution.

The audit probability has probably the clearest effect on tax evasion in the literature. However, the audit effect may interact with the fine rate, so we take this interaction into account in our

---

<sup>9</sup>Country and Study Fixed Effects are an absolute necessity as the countries and the studies in which the TEGs were completed vary in their characteristics. In total, 19 countries are covered in our sample, spanning around 840 million world inhabitants (or about 11% of world population). These countries include: Albania, Australia, Austria, Belgium, Canada, Colombia, The Czech Republic, Denmark, France, Germany, Israel, Italy, The Netherlands, Romania, Slovakia, Sweden, Taiwan, the United Kingdom, and the United States. In majority, the participants in these TEGs are *WEIRD* (or White, Educated, Industrialized, Rich, and Democratic), and they probably do not represent the social preferences of the world population (Henrich, Heine, and Norenzayan, 2010). In addition, summary statistics from Table 8 in the Appendix show important discrepancies between countries in term of average compliance behavior. For example, the mean compliance rate in the TEGs conducted in these countries is 91% in Albania and 34% in Slovakia; the overall country average is 65%. We can only call for running more standardized TEGs in more different continents, especially in Asia, South America, and Africa.

<sup>10</sup>From now on, unless otherwise stated, p-values systematically correspond to t-tests.

interpretation. To ease the understanding of this interaction, both continuous variables are thus centered. The impact of audit probability, when the fine rate is equal to its mean, is as expected positive and highly significant on the extensive margin of compliance. As a comparison, the impact of the fine rate is smaller, positive, and significant when the audit probability is equal to its mean. However, their interaction term is positive and highly significant on the extensive margin: when the fine rate increases, the magnitude of the impact of the audit probability also increases, and vice versa. This result is consistent with Blackwell (2010) in his previous meta-analysis.

With respect to the intensive margin of compliance, the audit probability has a strong, negative, and highly significant impact, even when the fine rate is equal to its mean. In contrast, the fine rate has a small, negative, and insignificant ( $p = 0.359$ ) impact when the audit probability is equal to its mean. Once again, the interaction term is highly significant, but it points toward the opposite direction this time.<sup>11</sup> Overall, the audit probability and fine rate compete to reduce tax evasion. In this case, when the audit probability increases and is coupled with an average fine, it leads to a decrease in tax compliance, so that evaders facing a higher probability of being checked will evade even more. Note that there is some experimental (and empirical) evidence that higher audit rates can sometimes backfire, leading to lower post-audit compliance (Alm, 2019).<sup>12</sup>

Despite their theoretical equivalence in the standard model of tax evasion, the audit probability and the fine rate have very different impacts on compliance. Their interaction terms also vary on the different margins: both variables can work together jointly or they can compete to reduce tax evasion. These results imply that each public policy tool must be used very carefully, depending on the context and the objectives of the tax administration.

Audits also have a greater deterrent impact than fines, again despite their theoretical equivalence. Indeed, the impact of fines is rather small. Further, the impact of both tools on the intensive margin is even counter-intuitive, a result that could be explained by persistent social norms within specific sub-samples of participants who decided to evade. For example, a norm of

---

<sup>11</sup>Note that this negative interaction holds mainly because of the presence of two countries that account for about 40 percent of the observations, or the USA and Italy (see Table 9 in the Appendix).

<sup>12</sup>There is also some evidence that higher audit rates can sometimes backfire, leading to lower post-audit compliance; that is, an audited individual may actually reduce his or her post-audit compliance, sometimes termed a “*bomb-crater effect*”. This effect is generally found in laboratory experiments in which tax compliance of audited taxpayers falls immediately after an audit (Mittone, 2006; Maciejovsky, Kirchler, and Schwarzenberger, 2007; Kastlunger, Kirchler, Mittone, and Pitters, 2009), before recovering somewhat in succeeding rounds. However, with some exceptions (Mendoza, Wielhouwer, and Kirchler, 2017), field data generally find little or no evidence of a bomb-crater effect (Erard, 1990). One explanation for the bomb-crater effect is that deterrence may “crowd out” an individual’s “intrinsic motivation” to pay taxes (Frey, 1997), and a similar finding has been found in field data for corporate taxpayers (DeBacker, Heim, Tran, and Yuskavage, 2015). Other explanations are that individuals may update their subjective audit probabilities following an audit, that they may attempt to “restore” their income following an audit, or the audit may alter their perceptions of the efficiency of the tax administration (Mittone, Panebianco, and Santoro, 2017). There is also some experimental work that finds that delayed feedback on tax audits is more effective in improving compliance than immediate feedback, possibly because delay leads individuals to overweight audit probabilities (Kogler, Mittone, and Kirchler, 2016). This issue remains unresolved.

non-compliance has been observed in Choo, Fonseca, and Myles (2015), where students comply less but also react differently to the occurrence of an audit.

Having a tax amnesty decreases the probability of being a full complier, and this effect is highly significant. The tax amnesty also has a negative impact on the intensive margin, although this effect is not significant ( $p = 0.070$ ). These results suggest that an amnesty is a questionable tool for improving tax compliance.

The tax rate has a negative and highly significant impact on the extensive margin. The impact is positive on the intensive margin, but very close to 0 and not significant ( $p = 0.836$ ). In these estimation results, the impact of tax rate is therefore unambiguous, despite the theoretical prediction by Yitzhaki (1974) that an increase in the tax rate should have a positive (and a somewhat counter-intuitive) impact on compliance.

A flat rate tax system is the only variable that has a negative and highly significant impact on both margins. It is thus clear cut that a flat tax system discourages compliance, compared to a system where richer individuals pay more taxes. The explanation often put forward, that a progressive tax system increases the tax rate for most individuals and thereby increases tax compliance (Heinemann and Kocher, 2013), is not supported here. We have seen above that tax rate has in fact a negative impact on compliance. We thus hypothesize that the perceived fairness of the tax system could be driving this result, as in Doerrenberg and Peichl (2013), where the authors also showed that tax evasion was less frequent in countries with a high level of tax progressivity.<sup>13</sup>

#### 4.2.2 Experimental design variables

We now turn to the multivariate impact of our experimental design variables (e.g., framing, the way to request compliance, the kind of task implemented, earning an income, redistribution of the taxes or investment in a real life public good). The results are reported in Table 2, Column 2 for the extensive margin and Column 5 for the intensive margin.

Implementing an experiment with loaded framing (i.e., with the use of tax related terms) has the expected positive sign on the extensive margin, but the result is not significant ( $p = 0.441$ ). On the intensive margin, the sign points in the opposite direction, a result that is also not significant ( $p = 0.064$ ). This result seems surprising, as framing an experiment is one way to achieve the “parallelism” often thought to be required to help achieve external validity by mirroring real life parameters (Torgler, 2002; Muehlbacher and Kirchler, 2016). However, participants behave the

---

<sup>13</sup>However, the relatively low number of studies implementing a progressive tax system might affect the generalizability of this conclusion. Some other variables such as redistribution of the tax fund could interact with the progressivity of the tax system. Table 13 in the Appendix shows that, when public policy and experimental variables are pooled, the effects of having a flat tax and redistribution exist, but their interaction is not very useful because there is only one dataset that implements a progressive tax system without redistribution. We can only suggest that there is need for more experiments with non-standard experimental settings, such as varying the progressivity along with the redistribution of the taxes.

Table 2: Multivariate regressions of compliance decisions on public policy variables, experimental design variables and individual-level variables

Variable	Extensive margin			Intensive margin		
	(1)	(2)	(3)	(4)	(5)	(6)
	Coef. (St. E.)	M. Eff. (St. E.)	M. Eff. (St. E.)	Coef. (St. E.)	Coef. (St. E.)	Coef. (St. E.)
Random audit	-19 (.158)			.09 (.054)		
Audit probability	1.48*** (.141)			-.34*** (.039)		
Fine size	.11** (.037)			-.008 (.009)		
Audit#Fine	.53*** (.132)			-.29*** (.037)		
Amnesty	-5.02*** (.064)			-.04 (.022)		
Flat tax	-.22** (.076)			-.12*** (.030)		
Tax rate	-.70*** (.100)			.008 (.039)		
Loaded framing		.03 (.048)			-.07 (.039)	
Directive way		.42*** (.089)			.10 (.087)	
Tax task		-.09 (.095)			.08 (.077)	
Earned income		.02 (.065)			-.06 (.048)	
Redistribution		.15*** (.023)			.08*** (.017)	
RL public good		.17*** (.028)			.09*** (.024)	
Age			4e-5 (.001)			.0001 (.001)
Male			-.03*** (.010)			-.15*** (.021)
Student			.05 (.035)			-.11* (.054)
Income			-.06*** (.008)			.006 (.012)
Risk averse (HL)			.02* (.010)			.02 (.020)
Constant	-1.22*** (.219)	—	—	.19* (.076)	.10 (.103)	.32*** (.079)
# Observations	185,425	155,997	29,420	94,999	84,550	6,908
# Clusters	12,112	9,700	1,549	8,878	6,839	1,075
Round FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	NO	YES	YES	NO
Study FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	NO	YES	YES	NO

**Note.** For the extensive margin estimations, we use a Probit model on the decision to evade (average marginal effect for Columns 2 and 3), where the dependent variable equals 1 if declared income is equal to earned income, to 0 otherwise. For the intensive margin estimations, we use OLS regressions on the compliance rate (defined as income declared divided by income earned or received), estimated on the subsample of evaders. Both models report in parentheses standard errors clustered at the individual level.  
Significance levels: \* : 5% \*\* : 1% \*\*\* : 0.1%

same when they take part in a TEG versus a neutral lying game. The results coming from neutral experiments thus seem valid in studying tax evasion behavior.

A directive way to ask for compliance has a highly significant and positive impact on the extensive margin. It has a positive but insignificant impact on the intensive margin ( $p = 0.247$ ). When participants decide to evade, they barely change their behavior whether the framing of the request to declare taxes is directive or not. To ease replicability of experimental results, this result argues for systematically reporting the way taxes are requested in the design of the experiment, because different requests may lead to significantly different results.

When participants are asked to declare taxes rather than income, the impact is not significant on both margins. Even though, no impact is found here, authors should likely focus on implementing a TEG with an income declaration task because this task is closer to what is happening in real life. Note that most TEGs are already implemented with an income declaration task.

Earning an income has a positive sign on the extensive margin of compliance, a result that is not significant ( $p = 0.733$ ). On the intensive margin, this variable has a negative but still insignificant effect ( $p = 0.177$ ). Whether participants' income is endowed or earned does not seem to matter for tax compliance decisions.

Redistribution of taxes that are collected to participants has a highly significant and positive impact on both the extensive and the intensive margins. Participants are thus sensitive to obtaining a personal benefit from their compliance. Investing collected taxes also has a highly significant and positive impact on both margins. Note that the magnitudes of the latter effects on both margins exceed the former ones; that is, participants are even more sensitive when a third party (e.g., a student or an environmental organization) benefits from their compliance.

### 4.2.3 Individual-level variables

Here, we investigate the multivariate impact of our individual level variables (age, gender, occupation, income and risk aversion). The results are reported in Table 2, Column 3 for the extensive margin and Column 6 for the intensive margin.

An additional year of age has the expected positive sign on the extensive margin, but its magnitude is extremely small and insignificant ( $p = 0.968$ ). Age has the same effect on the intensive margin of compliance ( $p = 0.931$ ). Older participants therefore do not seem to be more tax compliant, a result similar to Abeler, Nosenzo, and Raymond (2019).

Being a male has the expected negative and significant impact on the extensive margin ( $p = 0.001$ ), and it also has a negative and highly significant impact on the intensive margin ( $p < 0.001$ ). This result corroborates a well-documented effect: men are more prone to cheating.

Being a student has a positive but insignificant impact on the extensive margin of compliance ( $p = 0.143$ ). This result is likely due to the relatively low number of studies that have used non-students. However, a student subject pool has a negative and significant impact on the intensive margin of compliance, with a student having on average lower compliance by 11% compared to a

non-student, conditional on the decision to evade. This result is consistent with the finding that students are less compliant in TEGs (Görxhani and Schram, 2006; Choo, Fonseca, and Myles, 2015), it is somewhat contrary to the results of Alm, Bloomquist, and McKee (2015), although their focus was mainly on the comparative statics responses of students versus non-students, where they found similar effects.

The level of experimental income has a negative impact on the extensive margin, and it is highly significant. When income increases by one Euro, the probability of being a full complier decreases by 6%. Despite its sign being still negative, the level of income has no impact on the intensive margin of compliance ( $p = 0.583$ ). Tax evasion is thus increasing with the level of income, which echoes recent empirical evidence showing similar results (e.g., Johns and Slemrod, 2010; Alstadsæter, Johannesen, and Zucman, 2019).<sup>14</sup>

Risk aversion is generally captured via the Holt and Laury (2002) approach, in which participants are asked to choose between safe and risky lotteries. Being risk averse has a positive but very low impact on the extensive margin of compliance, which is significant at the 5% threshold ( $p = 0.014$ ). On the intensive margin, the effect is in the same direction, but it is not significant ( $p = 0.285$ ). This surprising absence of convincing result calls into question the use of Holt and Laury (2002) mechanism in TEGs. A more thorough study of risk aversion measures is discussed in our next section.<sup>15</sup>

### 4.3 Additional results on the meta-analysis

Last, we conduct a range of additional tests on the results from our meta-analysis.

#### 4.3.1 Audit rules

In Table 3 we report the results of estimations that examine in detail random audits versus endogenous audits, in part because the impact of a random audit was unclear in our main analysis. We compare the performance of different audit types, all of an endogenous (or strategic) nature (*Cordination*, *Risk-based*, *Cutoff*, and *Invest* audit rules), with random audits as the omitted category. The estimation results in Table 3 show variable effects for the endogenous audit rules, both on the extensive and the intensive margin. Only two audit rules, based on the investment of tax revenue from taxpayers and on the estimation of one’s level of tax evasion, perform significantly better than a random audit. Once again, it should be noted that these observations are highly unbalanced, with a vast majority of the observations corresponding to a random audit, and these observations are also quite heterogeneous. These results remain to be interpreted with caution,

---

<sup>14</sup>The impact of income does not change when we consider an income eliminating the differences in price levels between countries, see Table 10 in the Appendix where Income has been replaced by Income (PPP).

<sup>15</sup>Table 11 in the Appendix features the impact of risk attitude pooled with the public policy variables on lab tax compliance. The results still show a somewhat positive impact of risk aversion on both margins.

Table 3: Multivariate regressions of compliance decisions on audit rules

Variable	Extensive margin		Intensive margin	
	M. Eff.	(St. E.)	Coef.	(St. E.)
Coordination	.03	(.035)	-.08	(.053)
Cutoff	.0007	(.062)	-.03	(.054)
Invest	.21***	(.045)	.09*	(.041)
Risk-based	.52***	(.044)	.47***	(.046)
Constant	--	-	.30***	(.028)
# Observations	236,953		110,912	
# Clusters	14,928		10,195	
Round FE	YES		YES	
Country FE	YES		YES	
Study FE	YES		YES	
Year FE	YES		YES	

**Note.** For the extensive margin regression, we use Probit analysis (average marginal effects). For the intensive margin regression, we use OLS analysis, estimated on the subsample of evaders. The omitted category is a random audit. Both models report in parentheses standard errors clustered at the individual level.

Significance levels: \*: 5% \*\*: 1% \*\*\*: 0.1%

especially given the clear theoretical result that endogenous audit rules are better able to deter tax evasion than random audit rules (Malézieux, 2018).

#### 4.3.2 Real life public goods

We have seen in Table 2 that investing collected taxes in a real life public good has a positive impact on both margins of compliance. However, it is of interest to determine the effects of different real life public goods on laboratory tax compliance. As discussed earlier, these real life public goods can be of five different types: *Research*, *Different*, *Government*, *Organization*, and *Student*. These estimation results are reported in Table 4, where the omitted category is the Research public good. The estimation results show no differences, both on the extensive and the intensive margin. It does not seem to matter what kind of organization is used in a TEG, and a *Research* public good seems like a representative type of real life public good to implement in a TEG.

#### 4.3.3 Risk attitude

Aside from the Holt and Laury (2002) measure, numerous risk attitude measures have also been used in the literature. Bazart and Bonein (2014) use 20 lotteries and risk averse individuals are those choosing more than 5 safe lotteries. We denote this variable *20 Lotteries* or *20L*. Doerrenberg and Duncan (2014a) use only one lottery (asking subjects to choose between a certain payoff of \$50 and a gamble that pays \$100 with a probability of 0.5) and risk averse individuals are those choosing the sure payment over the lottery (*1 Lottery* or *1L*). Some researchers have also used tasks

Table 4: Multivariate regressions of compliance decisions on real life public good destinations

Variable	Extensive margin		Intensive margin	
	M. Eff.	(St. E.)	Coef.	(St. E.)
Different	.01	(.074)	-.11	(.099)
Government	-.08	(.103)	-.05	(.075)
Organization	.01	(.097)	-.01	(.079)
Student	.08	(.058)	-.08	(.116)
Constant	--	-	.28**	(.092)
# Observations	17,200		9,827	
# Clusters	5,523		2,505	
Round FE	YES		YES	
Country FE	YES		YES	
Study FE	YES		YES	
Year FE	YES		YES	

**Note.** For the extensive margin regression, we use Probit analysis (average marginal effects). For the intensive margin regression, we use OLS analysis, estimated on the subsample of evaders. The omitted category is a research public good. Both models report in parentheses standard errors clustered at the individual level. Significance levels: \*: 5% \*\*: 1% \*\*\*: 0.1%

other than lotteries to capture risk attitudes. Coricelli, Joffily, Montmarquette, and Villeval (2010) use a portfolio choice of how much to invest in a risky asset, following Charness and Gneezy (2010), which we call *Portfolio Choice* or *PC*, with risk averse participants those investing less than half of their endowment. Mittone and Saredi (2016) use a bomb risk elicitation task where participants have to decide how many boxes to pick to earn money, risking to pick one box containing a bomb (Crosetto and Filippin, 2013); we call this variable *Bomb Risk Elicitation Task* or *BRET*, where risk averse participants are those picking less than 50 boxes. Other researchers have opted for subjective questions to let participants assess their own level of risk attitude. Several researchers ask the question: “*Are you generally a person who is completely willing to take risks, or do you normally try to avoid taking them?*”, ranging from 1 (unwilling to take risks) to 10 (completely willing to take risks) (Ottone, Ponzano, and Andrighetto, 2018; Zhang, Andrighetto, Ottone, Ponzano, and Steinmo, 2016; D’Attoma, Volintiru, and Steinmo, 2017; D’Attoma, Volintiru, and Malézieux, 2020; D’Attoma, 2018). We call this variable *Subjective Question (1)* or *SQ1*. Casal, Kogler, Mittone, and Kirchler (2016) also ask one item to participants: “*Generally speaking, would you characterize yourself as someone who is willing to take risks, or as someone who is avoiding risks?*”, ranging from 1 (absolute risk aversion) to 9 (absolute risk seeking). We call this variable *Subjective Question (2)* or *SQ2*. One item from Lefebvre, Pestieau, Riedl, and Villeval (2015) asks “*Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?*”, ranging between 1 (risk averse) and 9 (willing to take full risks). We call this variable *Subjective Question (3)* or *SQ3*. In all of these questionnaires, a rating of 5 or less corresponded to risk

Table 5: Pairwise correlation coefficients between risk attitude measures and tax compliance

	Compliance	Holt & Laury
Holt and Laury (2002)	-.02 (42,041)	-
Subjective question (1)	.17 (24,640)	-
Subjective question (3)	-	.26 (7,710)
20 Lotteries	.10 (5,760)	-
Portfolio Choice	.16 (1,440)	-
1 Lottery	.085 (1,719)	-
Subjective question (2)	.20 (492)	-
Dospert-Financial	.18 (450)	.18 (450)

**Note.** For the first column, we report pairwise correlation coefficients between compliance and risk attitude measures. For the second column, we report pairwise correlation coefficients between the Holt and Laury (2002) number of safe choices with risk attitude measures. The number of observations is in parentheses. All the correlations are significant at:  $p < 0.0001$ .

aversion. Finally, questionnaires can also be completed by participants, such as the DOSPERT Scale from Blais (2006) (*Dospert-Financial* or *Dospert*). The financial subscale of the DOSPERT (6 items) has been filled by participants in Bühren and Kundt (2014). Risk averse individuals are those scoring less than 24 points at this subscale.<sup>16</sup>

We use this information to report in Table 5 (left part) pairwise correlations between the compliance rate and the different risk attitude measures. For ease of interpretation, all the risk attitude measures are recoded so that the higher the score the higher is risk aversion, which implies that we should see a positive correlation between risk aversion and tax compliance. For the most part, these results show that the risk attitude measures correlate significantly ( $p < 0.0001$ ) with compliance (which is probably essentially due to the very large sample studied here). Surprisingly, the number of safe choices for the Holt and Laury (2002) measure is negatively correlated with tax compliance, with a coefficient almost 0. All the other risk attitude measures have a positive correlation, but they vary in their intensity. The weakest correlation is with the Doerrenberg and Duncan (2014a) *1 Lottery* measure ( $r = 0.08$ ); the strongest correlation is with the Casal, Kogler, Mittone, and Kirchler (2016) one item question ( $r = 0.20$ ), *Subjective Question (2)*. Regardless,

<sup>16</sup>The information above is summarized in Table 12 in the Appendix. Note that the average number of observations that could be considered as risk averse varies widely across methodologies, ranging from 27% in SQ1 and PC to 87% in Dospert.

according to Mukaka (2012), all the present correlations can be judged as negligible. Replacing the variable of interest by one capturing the extensive margin of compliance or the risk attitude measures by a classification in terms of risk aversion does not change any of those conclusions.

The most used measure of risk aversion in our dataset ( $N = 42,041$ ), or *Holt & Laury*, is also the one with the lowest correlation with tax compliance. We run an additional analysis to study if Holt and Laury (2002) lotteries are consistent with other measures of risk attitudes. Bühren and Kundt (2014) and Lefebvre, Pestieau, Riedl, and Villeval (2015) paired the Holt and Laury (2002) measure with the financial subscale of the DOSPRT and a one item question. Table 5 (right part) shows pairwise correlations between the Holt and Laury (2002) measure and *Dospert-Financial* and a one item question (*Subjective Question (3)*). The results show expected positive correlations, but once again they are low and almost negligible in size.<sup>17</sup>

In conclusion, no risk measure seems particularly satisfactory when it comes to correlation with tax compliance.<sup>18</sup> As in Menkhoff and Sakha (2017), risk measures are positively correlated but perform heterogeneously. Despite being the most common, the Holt and Laury (2002) measure, along with other lottery-based measures, especially perform worse. This result could be driven by the difficulties in understanding the Holt and Laury (2002) task, such as a lack of mathematical ability or an unfamiliarity with probabilities (Charness, Gneezy, and Imas, 2013; Grüner, 2017). Zhou and Hey (2018) also argue that risk measures are only pertinent in corresponding contexts. As a TEG is more than a simple amoral choice under risk (Baldry, 1987), it may not correlate well with lotteries for this reason. The lack of convincing correlations between risk aversion measures and tax evasion behavior could also find its explanation within the prospect theory framework of Kahneman and Tversky (1979). Indeed, most of the risk elicitation procedures used in the

---

<sup>17</sup>See Pedroni, Frey, Bruhin, Dutilh, Hertwig, and Rieskamp (2017) or Frey, Pedroni, Mata, Rieskamp, and Hertwig (2017) for an in-depth discussion of inconsistency between risk preference measures.

<sup>18</sup>The treatment of risk attitudes yields different general results in both laboratory and field experiments (see e.g. Friedman, Isaac, James, and Sunder, 2014; Galizzi, Machado, and Miniaci, 2016; Charness, Garcia, Offerman, and Villeval, 2019). It poses challenges in experimental economics, for several reasons. First, because of the relatively small stakes used in most experiments, risk aversion is inconsistent with expected utility theory, as demonstrated by Rabin (2000), who concludes that “...within the expected-utility model, anything but virtual risk neutrality over modest stakes implies manifestly unrealistic risk aversion over large stakes” (p. 1281-1282). Second, from the perspective of behavioral economics, testing for risk aversion is problematic because, relative to the reference point, many individuals are risk-averse or risk-neutral in gains but risk-seeking in losses (Kahneman and Tversky, 1979). As a result, risk preference may depend on whether decisions are framed as gains or losses, and methods for testing risk aversion do not typically test for framing effects. Third, the experiments that we examine involve many rounds, and risk preferences may change over rounds since repeated market experience can induce framing effects (Coursey, Hovis, and Schulze, 1987; List, 2003). However, methods for testing risk aversion do not typically consider that risk preferences might change with experience. Finally and relatedly, these methods do not consider how risk preferences might change with, say, framing effects or with the weighting of high and low probabilities. For all of these reasons, a single measure of risk aversion is not likely to be a useful concept in the experimental studies that we examine because these experiments typically have modest stakes, include both gains and losses, employ multiple rounds with changing subject experience, and use probabilities of audit that imply different weights relative to actual probabilities of audit.

present studies focus on decisions in the gain domain while the prospect of paying a fine is located in the loss domain. Another avenue for research on tax evasion in the lab could be to correlate declarations with a measure of loss aversion, such as in Tanaka, Camerer, and Nguyen (2010) or in Von Gaudecker, Van Soest, and Wengstrom (2011). To the best of our knowledge, comparing tax evasion behavior with a measure of loss aversion has not been implemented in the lab before. As a result, we recommend ignoring risk measures, using many risk measures (Frey, Pedroni, Mata, Rieskamp, and Hertwig, 2017), or even continuing a search for a single measure that is adapted to the specific tax evasion context (Zhou and Hey, 2018). A one item question, such as in Casal, Kogler, Mittone, and Kirchler (2016), could well be the best option.

#### 4.3.4 Specification on behavioral types

The following analysis sorts people according to their behavioral type. In the literature, it is often found that a minority of people lies to the full extent, another minority never lies, and the majority varies between lie and truth (Mazar, Amir, and Ariely, 2008; Ariely and Jones, 2012; Fischbacher and Föllmi-Heusi, 2013) and the TEG is no exception to this rule (Jacquemet, Luchini, Malézieux, and Shogren, 2020). In the present dataset, about 3.4% of participants are consistent evaders, 13.2% are consistent compliers and 83.2% varies in between. The specification in Table 14 presents the estimation results of a multinomial logistic regression where a new variable is used: participants' behavioral type (0= full evader, 1= people who varies between both and 2=full complier), the base being the inconsistent participants. The same variables as in Table 2 are analyzed here.

The fact of implementing a random audit is associated with an increase in the relative log odds of being a full evader compared to being inconsistent. As expected, audit probability, fine size and their interaction have coefficients associated with a decrease in the relative log odds of being a full evader. Both the flat tax system ( $p = 0.01$ ) and the tax rate have a counter intuitive impact on the relative log odds of being a full evader. The results are similar when considering the impact on the relative log odds of being a full complier vs. inconsistent, especially when focusing on the remaining significant variables.

Implementing a loaded framing goes in the expected direction in terms of compliance. It has a negative impact on the relative log odds of being a full evader and a positive one on being a full complier. Having a directive way to ask for compliance in the instructions is associated with an increase in the relative log odds of being a full complier with respect to being inconsistent. Implementing a tax task has also a positive impact on the relative log odds of being a full complier vs. inconsistent ( $p = 0.049$ ). Earning an income instead of giving it has a positive impact on the relative log odds of being both a full evader and a full complier. Redistributing taxes collected to participants and investing it in a real life public good both go in the expected direction, with a negative impact on the relative log odds of being a full evader and a positive one on being a full complier (at the exception of RL public good which is significant at  $p = 0.015$  on the odds of

being a full complier).

Only the fact of being a male has a significant impact when considering individual-level variables. Being a male is associated with a 1.29 increase in the relative log odds of being a full evader, vs. an inconsistent ( $p = 0.019$ ), and a 0.28 decrease in the in the relative log odds of being a full complier ( $p = 0.019$ ).

#### 4.3.5 Robustness checks

Here we present different robustness checks from alternative models and specifications.

Table 13 in Appendix pools the public policy and experimental design variables to compare their impacts on both the extensive and intensive margins. With the exception of Income, the other individual-level variables (e.g., age, gender, occupation, risk aversion) have to be excluded due to sample size and collinearity problems. Despite losing about 2/3 of the total observations and about 1/2 of the total individuals, the results presented here are fairly robust: few variables have a different impact when considered pooled.

Table 15 in Appendix reproduces exactly Table 2, except that the Probit model is replaced by a Logit model and the OLS model is replaced by a GLM model. The results show almost no differences between both settings.

Table 16 in Appendix presents the average effect of public policy, experimental design, and individual-level variables on tax compliance, without distinguishing between the extensive and the intensive margins of compliance. Columns 1, 2 and 3 depict a Tobit model; Columns 4, 5, and 6 present OLS regressions. The results show highly similar estimation outputs for both models. However, compared to Table 2, it seems that the negative interaction between the audit probability and the fine rate prevails on the whole sample. Both tools, audits and fines, seem to compete with each other in fighting tax evasion.

Table 17 in Appendix reproduces Table 2, except that standard errors are here clustered at the study level. Fewer variables are now significant, but most has the same qualitative impacts.

## 5 Conclusion

Our meta-analysis has collected the largest number of observations originating from Tax Evasion Games (TEGs), more than any previous study, in order to estimate the impact of public policy variables, experimental design variables, and individual-level variables on tax compliance.

Regarding public policy variables, our results show that audits and fines interact differently on the extensive and intensive margins. These variables have a positive impact on the extensive margin, and their interaction is positive: the higher the audit probability, the greater is the impact of the fine rate on the decision to comply, and vice versa. However, their interaction is negative on the intensive margin: the higher the audit probability, the more negative is the impact of the fine on tax compliance coming from evaders, and vice versa. In addition, a flat tax system, the

tax rate, and a tax amnesty have unambiguously negative impacts on tax compliance. Finally, we find that two endogenous audit rules achieve higher compliance than a random audit rule, while other endogenous audit rules have no significant impact.

These results underscore the imperfect substitutability of fines and audits, with a much stronger impact of the latter than the former. Even so, these public policy variables are complementary tools that need to be combined efficiently. A system in which taxpayers have few chances to be audited but severely fined in case of evasion could be the easiest to implement, on top of being rather costless for the tax administration. However, their impacts may prove to be small or even to backfire (as found on the intensive margin). In this last case, and if such an effect was confirmed with real life data, a productive move for tax administrations may well involve a move from a “*cops-and-robbers*” to a “*service-and-client*” perspective (Kirchler, 2007), to a “trust” perspective (Alm and Torgler, 2011), or to an emphasis on understanding taxpayers’ tax morale (Torgler, 2002; Luttmer and Singhal, 2014).

Our results also demonstrate that a tax system should not implement a tax amnesty: when taxpayers are not punished for their evasion, it decreases their probability of fully complying. We also find that a tax system should have a lower tax rate to raise compliance and that the tax system has to be seen as fair and progressive to encourage compliance.

The present meta-analysis provides strong support for the positive impact on tax compliance of audit probability and fine size, as predicted by Allingham and Sandmo (1972). However, we find this result only on the extensive margin, where audits and fines work hand in hand to make taxpayers shift to full compliance. In contrast, we find different results on the intensive margin, where audits and fines work against each other. Also, Allingham and Sandmo (1972) predict that the tax rate has an ambiguous impact on tax compliance when the penalty is imposed on evaded income, while Yitzhaki (1974) shows that the tax rate has a clear positive impact when the penalty is imposed on evaded taxes. Our meta-analysis finds evidence contrary to the prediction of Yitzhaki (1974), at least on the extensive margin.

With respect to the experimental design variables, we find that several features have a positive impact on the extensive margin, such as asking for compliance in a directive way, redistributing taxes that are collected to the participants, and investing those taxes in a real life public good. Only the redistribution of taxes to participants, and the investment of taxes in a real life public good have a positive and significant impact on the intensive margin of compliance. Using loaded framing, implementing a tax task and making participants earn an income do not seem to matter at all.

These experimental variable results suggest some recommendations on how best to design a TEG. It has been consistently recommended to run a TEG with earned income and a tax frame (Torgler, 2002; Muehlbacher and Kirchler, 2016). Our results suggest that these design features do not affect the experimental results; however, their inclusion likely increases the parallelism needed to generalize from TEGs to the real world. Relatedly, we believe that a directive way to

ask for compliance should be implemented, as this features also mirrors real life. Regardless, we believe that it is important for researchers to systematically report the method used for framing their experiment. In addition, TEGs should ask participants to declare income rather than taxes, since a tax task does not correspond to what is happening in real life. Finally, as emphasized by Muehlbacher and Kirchler (2016), we recommend that taxes that are collected be invested either in a group fund redistributed to the participants, or be given as a donation for an organization.

Our results finally suggest that individual-level variables matter for tax compliance behavior. We find that the typical profile of the tax evader in the laboratory is a male and someone with a higher income who is risk averse. Being a student increases his share of evasion. Surprisingly, the age of the subject does not seem to matter.

The experimental literatures on general cheating and tax evasion have largely ignored each other. This meta-analysis is also the opportunity to link both. In the present paper, we have learned that introducing a TEG in a framed or in a neutral way did not change the generalizability of the results obtained, along with changing the task from an income to a tax declaration. In many respects, our results are broadly similar to the results of Abeler, Nosenzo, and Raymond (2019) in their meta-analysis of truth-telling, especially in the effects of individual characteristics on cheating on taxes in experiments: age having no impact along with men and students increasing cheating. Participants in TEG compared to lying experiments generally do not cheat as much as they could and are much more honest than what the standard economic theory predicts. Audit probability having one of the strongest impacts – higher than fine size – also tells thus that being perceived as honest matters for tax compliance. One simple explanation could not explain such heterogeneity in compliance, but as with general cheating experiments, tax evasion could be a matter of personal proportional lying costs along with a preference for being seen as honest. More research would be needed to confirm it further.

We should acknowledge various limitations of our meta-analysis. First, the application of laboratory experiments is far from homogeneous. Indeed, the TEGs collected here have been run by researchers from economics, psychology, accounting, and marketing backgrounds, which may lead to some subtle differences in experimental design. Even so, the specific background of authors has generally been found to have little impact on the behavior of participants in a meta-analysis (Balliet, Mulder, and Van Lange, 2011; Abeler, Nosenzo, and Raymond, 2019).

Second, we were not able to integrate all of the data from all of the TEGs. Over the 40 years of laboratory experiments, numerous datasets have been lost, especially for the papers published during the first 30 years. Even so, it should be noted that, despite reaching only 50% of all of the articles published/written since 1978, we were able to recover above 76% of the articles published/written since 2008. Further, there might be some publication (or “file-drawer”) bias (Borenstein, Hedges, Higgins, and Rothstein, 2011). However, we found few datasets from working papers that authors did not wish to share with us ( $N = 4$ ). Running an economic experiment is expensive in terms of time and money, so there is clearly an incentive to publish any finding. We

doubt that the size of the file-drawer problem is important enough to change our main conclusions.

Third, nearly all of the lab experiments collected here are run with fiscal parameters (especially the audit probability) known with certainty by the participants. In real life, this is often not the case. More evidence is needed to understand the effects of ambiguity.

Finally, some of our findings are inconclusive and surprising. The finding that an endogenous audit rule does not in general perform better than a random audit rule should be viewed with some caution, especially given the common theoretical result that endogenous audit rules are better able to deter tax evasion than random audit rules (Malézieux, 2018). The lack of correlation between standard risk attitude measures like Holt and Laury (2002) and tax compliance, is also surprising. Last, few TEGs are run in different countries and those countries are highly similar. Important areas for further investigation include such issues as studying the performance of different audit types, creating new tools to measure risk aversion in the tax domain, conducting new experiments to test between motivations for lying and running more standard and non-standard TEGs with larger numbers of subjects in different institutional and country characteristics.

## References

- ABELER, J., D. NOSENZO, AND C. RAYMOND (2019): "Preferences for truth-telling," *Econometrica*, 87(4), 1115–1153.
- ABRAHAM, M., K. LOREK, F. RICHTER, AND M. WREDE (2017): "Collusive tax evasion and social norms," *International Tax and Public Finance*, 24(2), 179–197.
- ADRES, E., D. R. VASHDI, AND Y. ZALMANOVITCH (2016): "Globalization and the retreat of citizen participation in collective action: A challenge for public administration," *Public Administration Review*, 76(1), 142–152.
- ALLINGHAM, M. G., AND A. SANDMO (1972): "Income tax evasion: a theoretical analysis," *Journal of Public Economics*, 1, 323–338.
- ALM, J. (1988): "Uncertain tax policies, individual behavior, and welfare," *The American Economic Review*, 78(1), 237–245.
- (2014): "Does an uncertain tax system encourage "aggressive tax planning"?", *Economic Analysis and Policy*, 44(1), 30–38.
- (2019): "What motivates tax compliance?," *Journal of Economic Surveys*, 33(2), 353–388.
- 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): "Institutional uncertainty and taxpayer compliance," *The American Economic Review*, 82(4), 1018–1026.
- ALM, J., G. H. MCCLELLAND, AND W. D. SCHULZE (1992): "Why do people pay taxes?," *Journal of Public Economics*, 48(1), 21–38.
- ALM, J., G. H. MCCLELLAND, AND W. D. SCHULZE (1999): "Changing the social norm of tax compliance by voting," *Kyklos*, 52(2), 141–171.
- ALM, J., N. MCKEE, AND M. BECK (1990): "Amazing grace: Tax amnesties and compliance," *National Tax Journal*, 43(1), 23–37.
- ALM, J., I. SANCHEZ, AND A. DE JUAN (1995): "Economic and noneconomic factors in tax compliance," *Kyklos*, 48(1), 3–18.
- ALM, J., AND B. TORGLER (2011): "Do ethics matter? Tax compliance and morality," *Journal of Business Ethics*, 101(4), 635–651.
- ALSTADSÆTER, A., N. JOHANNESSEN, AND G. ZUCMAN (2019): "Tax evasion and inequality," *The American Economic Review*, 109(6), 2073–2103.
- ANDREONI, J., B. ERARD, AND J. FEINSTEIN (1998): "Tax compliance," *The Journal of Economic Literature*, 36(2), 818–860.
- ANDRIGHETTO, G., N. ZHANG, S. OTTONE, F. PONZANO, J. D'ATTOMA, AND S. STEINMO (2016): "Are some countries more honest than others? Evidence from a tax compliance experiment in Sweden and Italy," *Frontiers in Psychology*, 7, 472.
- ARIELY, D., AND S. JONES (2012): *The (honest) truth about dishonesty*. New York, NY: Harper Collins Publishers.

- BALDRY, J. C. (1986): "Tax evasion is not a gamble: A report on two experiments," *Economics Letters*, 22(4), 333–335.
- BALDRY, J. C. (1987): "Income tax evasion and the tax schedule: some experimental results," *Public Finance*, 42(3), 357–383.
- BALLIET, D., L. B. MULDER, AND P. A. VAN LANGE (2011): "Reward, punishment, and cooperation: A meta-analysis," *Psychological Bulletin*, 137(4), 594–615.
- BAYER, R. C., AND M. SUTTER (2009): "The excess burden of tax evasion – An experimental detection-concealment contest.," *European Economic Review*, 53(5), 527–543.
- BAZART, C., AND A. BONEIN (2014): "Reciprocal relationships in tax compliance decisions," *Journal of Economic Psychology*, 40, 83–102.
- BECKER, G. S. (1968): "Crime and punishment: An economic approach," *Journal of Political Economy*, 76(2), 169–217.
- BECKER, W., H.-J. BÜCHNER, AND S. SLEEKING (1987): "The impact of public transfer expenditures on tax evasion: An experimental approach," *Journal of Public Economics*, 34(2), 243–252.
- BEER, S., M. KASPER, AND J. LOEPRICK (2016): "Puzzling tax law–Behavioral responses to complexity," Working Paper.
- BERNASCONI, M., L. CORAZZINI, AND R. SERI (2014): "Reference dependent preferences, hedonic adaptation and tax evasion: Does the tax burden matter?," *Journal of Economic Psychology*, 40, 103–118.
- BLACKWELL, C. (2010): "A meta-analysis of incentive effects in tax compliance experiments," in *Developing Alternative Frameworks for Explaining Tax Compliance*, ed. by J. Alm, J. Martinez-Vazquez, and B. Torgler, pp. 97–112. New York, NY: Routledge Publishing.
- BLACKWELL, C., AND M. MCKEE (2012): "Alchian and Allen visit the IRS: Costly audits and taxpayer compliance," *Applied Economics Letters*, 19(17), 1731–1734.
- BLAIS, A.-R. (2006): "A domain-specific risk-taking (DOSPERT) scale for adult populations," *Judgment and Decision Making*, 1(1), 33–47.
- BLAUFUS, K., M. BRAUNE, J. HUNDSDOERFER, AND M. JACOB (2015): "Self-serving bias and tax morale," *Economics Letters*, 131, 91–93.
- BORENSTEIN, M., L. V. HEDGES, J. P. HIGGINS, AND H. R. ROTHSTEIN (2011): *Introduction to Meta-analysis*. Hoboken, NJ: John Wiley & Sons.
- BORGHANS, L., J. J. HECKMAN, B. H. GOLSTEYN, AND H. MEIJERS (2009): "Gender differences in risk aversion and ambiguity aversion," *Journal of the European Economic Association*, 7(2-3), 649–658.
- BOSCO, L., AND L. MITTONE (1997): "Tax evasion and moral constraints: some experimental evidence," *Kyklos*, 50(3), 297–324.
- BOYLAN, S. J. (2010): "Prior audits and taxpayer compliance: Experimental evidence on the effect of earned versus endowed income," *Journal of the American Taxation Association*, 32(2), 73–88.
- BROCKMANN, H., P. GENSCHEL, AND L. SEELKOPF (2016): "Happy taxation: Increasing tax compliance through positive rewards?," *Journal of Public Policy*, 36(3), 381–406.

- BRUNER, D. M., J. D'ATTOMA, AND S. STEINMO (2017): "The role of gender in the provision of public goods through tax compliance," *Journal of Behavioral and Experimental Economics*, 71, 45–55.
- BRUTTEL, L., AND T. FRIEHE (2014): "On the path dependence of tax compliance," *European Economic Review*, 65, 90–107.
- BUCKENMAIER, J., E. DIMANT, AND L. MITTONE (2018): "Effects of institutional history and leniency on collusive corruption and tax evasion," *Journal of Economic Behavior & Organization*.
- BÜHREN, C., AND T. C. KUNDT (2014): "Does the level of work effort influence tax evasion? Experimental evidence," *Review of Economics*, 65(2), 137–158.
- CADSBY, C. B., E. MAYNES, AND V. U. TRIVEDI (2006): "Tax compliance and obedience to authority at home and in the lab: A new experimental approach," *Experimental Economics*, 9(4), 343–359.
- CASAL, S., C. KOGLER, L. MITTONE, AND E. KIRCHLER (2016): "Tax compliance depends on voice of taxpayers," *Journal of Economic Psychology*, 56, 141–150.
- CASAL, S., AND L. MITTONE (2016): "Social esteem versus social stigma: The role of anonymity in an income reporting game," *Journal of Economic Behavior & Organization*, 124, 55–66.
- CASTRO, M. F., AND I. RIZZO (2014): "Tax compliance under horizontal and vertical equity conditions: An experimental approach," *International Tax and Public Finance*, 21(4), 560–577.
- CHARNESS, G., T. GARCIA, T. OFFERMAN, AND M. C. VILLEVAL (2019): "Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory?," Working Paper.
- CHARNESS, G., AND U. GNEEZY (2010): "Portfolio choice and risk attitudes: An experiment," *Economic Inquiry*, 48(1), 133–146.
- CHARNESS, G., U. GNEEZY, AND A. IMAS (2013): "Experimental methods: Eliciting risk preferences," *Journal of Economic Behavior & Organization*, 87, 43–51.
- CHOO, C. L., M. A. FONSECA, AND G. D. MYLES (2015): "Do students behave like real taxpayers in the lab? Evidence from a real effort tax compliance experiment," *Journal of Economic Behavior & Organization*, 124, 102–114.
- CORICELI, G., M. JOFFILY, C. MONTMARQUETTE, AND M.-C. VILLEVAL (2010): "Cheating, emotions, and rationality: An experiment on tax evasion," *Experimental Economics*, 13(2), 226–247.
- CORICELI, G., E. RUSCONI, AND M.-C. VILLEVAL (2014): "Tax evasion and emotions: An empirical test of re-integrative shaming theory," *Journal of Economic Psychology*, 40, 49–61.
- COURSEY, D. L., J. L. HOVIS, AND W. D. SCHULZE (1987): "The disparity between willingness to accept and willingness to pay measures of value," *The Quarterly Journal of Economics*, 102(3), 679–690.
- CROSETTO, P., AND A. FILIPPIN (2013): "The "bomb" risk elicitation task," *Journal of Risk and Uncertainty*, 47(1), 31–65.
- DAI, Z. (2016): "Endogenous Crackdowns: Theory and Experimental Evidence," Working Paper.
- D'ATTOMA, J. (2018): "What explains the North – South divide in Italian tax compliance? An experimental analysis," *Acta Politica*, pp. 1–20.

- D'ATTOMA, J., C. VOLINTIRU, AND A. MALÉZIEUX (2020): "Gender, social value orientation, and tax compliance," *CEifo Economic Studies*, forthcoming.
- D'ATTOMA, J., C. VOLINTIRU, AND S. STEINMO (2017): "Willing to share? Tax compliance and gender in Europe and America," *Research & Politics*, 4(2).
- DAVIS, J., K. MENGERSEN, S. BENNETT, AND L. MAZEROLLE (2014): "Viewing systematic reviews and meta-analysis in social research through different lenses," *SpringerPlus*, 3(1).
- DEBACKER, J., B. T. HEIM, A. TRAN, AND A. YUSKAVAGE (2015): "Legal enforcement and corporate behavior: An analysis of tax aggressiveness after an audit," *The Journal of Law and Economics*, 58(2), 291–324.
- DOERRENBERG, P. (2015): "Does the use of tax revenue matter for tax compliance behavior?," *Economics Letters*, 128, 30–34.
- DOERRENBERG, P., AND D. DUNCAN (2014a): "Distributional implications of tax evasion: Evidence from the lab," *Public Finance Review*, 42(6), 720–744.
- DOERRENBERG, P., AND D. DUNCAN (2014b): "Experimental evidence on the relationship between tax evasion opportunities and labor supply," *European Economic Review*, 68, 48–70.
- DOERRENBERG, P., D. DUNCAN, AND C. ZEPPENFELD (2015): "Circumstantial risk: Impact of future tax evasion and labor supply opportunities on risk exposure," *Journal of Economic Behavior & Organization*, 109, 85–100.
- DOERRENBERG, P., AND A. PEICHL (2013): "Progressive taxation and tax morale," *Public Choice*, 155(3-4), 293–316.
- DULLECK, U., J. FOOKEN, C. NEWTON, A. RISTL, M. SCHAFFNER, AND B. TORGLER (2016): "Tax compliance and psychic costs: Behavioral experimental evidence using a physiological marker," *Journal of Public Economics*, 134, 9–18.
- DURHAM, Y., T. S. MANLY, AND C. RITSEMA (2014): "The effects of income source, context, and income level on tax compliance decisions in a dynamic experiment," *Journal of Economic Psychology*, 40, 220–233.
- EISENHAUER, J. G., D. GEIDE-STEVENSON, AND D. L. FERRO (2011): "Experimental Estimates of taxpayer ethics," *Review of Social Economy*, 69(1), 29–53.
- ENGEL, C. (2011): "Dictator games: A meta study," *Experimental Economics*, 14(4), 583–610.
- ERARD, B. (1990): "The influence of tax audits on reporting behavior," in *Why People Pay Taxes: Tax Compliance and Enforcement*, ed. by J. Slemrod, pp. 95–114. Ann Arbor, MI: The University of Michigan Press.
- ETCHART-VINCENT, N., AND E. TAUGOURDEAU (2018): "The influence of tax morality and others' tax evasion on one's fiscal behaviour," Working Paper.
- FISAR, M., T. REGGIANI, AND J. ŠPALEK (2018): "'There is good news!' Media negativity bias and tax compliance," Working Paper.
- 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.
- FORTIN, B., G. LACROIX, AND M.-C. VILLEVAL (2007): "Tax evasion and social interactions," *Journal of Public Economics*, 91(11-12), 2089–2112.

- FREY, B. (1997): *Not Just For the Money – An Economic Theory of Personal Motivation*. Cheltenham: Edward Elgar Publishing Limited.
- FREY, R., A. PEDRONI, R. MATA, J. RIESKAMP, AND R. HERTWIG (2017): “Risk preference shares the psychometric structure of major psychological traits,” *Science Advances*, 3(10).
- FRIEDLAND, N. (1982): “A Note on Tax Evasion as a Function of the Quality of Information about the Magnitude and Credibility of Threatened Fines: Some Preliminary Research,” *Journal of Applied Social Psychology*, 12(1), 54–59.
- FRIEDLAND, N., S. MAITAL, AND A. RUTENBERG (1978): “A simulation study of income taxation,” *Journal of Public Economics*, 10(1), 107–116.
- FRIEDMAN, D., R. M. ISAAC, D. JAMES, AND S. SUNDER (2014): *Risky curves: On the empirical failure of expected utility*. New York, NY: Routledge.
- GALIZZI, M. M., S. R. MACHADO, AND R. MINIACI (2016): “Temporal stability, cross-validity, and external validity of risk preferences measures: Experimental evidence from a UK representative sample,” Working Paper.
- GËRKHANI, K., AND A. SCHRAM (2006): “Tax evasion and income source: A comparative experimental study,” *Journal of Economic Psychology*, 27(3), 402–422.
- GLASS, G. V. (1976): “Primary, secondary, and meta-analysis of research,” *Educational Researcher*, 5(10), 3–8.
- GROSCH, K., AND H. A. RAU (2017): “Gender differences in honesty: The role of social value orientation,” *Journal of Economic Psychology*, 62, 258–267.
- GRÜNER, S. (2017): “Correlates of multiple switching in the Holt and Laury procedure,” *Economics Bulletin*, 37(1), 297–304.
- GUERRA, A., AND B. HARRINGTON (2018): “Attitude–behavior consistency in tax compliance: A cross-national comparison,” *Journal of Economic Behavior & Organization*, 156, 184–205.
- GUTH, W., AND R. SAUSGRUBER (2008): “Voting between tax regimes to fund a public good,” *Economics of Governance*, 9(4), 287–303.
- HARTL, B., E. HOFMANN, K. GANGL, M. HARTNER-TIEFENTHALER, AND E. KIRCHLER (2015): “Does the sole description of a tax authority affect tax evasion? The impact of described coercive and legitimate power,” *PloS one*, 10(4).
- HASHIMZADE, N., G. D. MYLES, AND B. TRAN-NAM (2013): “Applications of behavioural economics to tax evasion,” *Journal of Economic Surveys*, 27(5), 941–977.
- HEINEMANN, F., AND M. G. KOCHER (2013): “Tax compliance under tax regime changes,” *International Tax and Public Finance*, 20(2), 225–246.
- HENRICH, J., S. J. HEINE, AND A. NORENZAYAN (2010): “The weirdest people in the world?,” *Behavioral and Brain Sciences*, 33(2-3), 61–83.
- HOLT, C. A., AND S. K. LAURY (2002): “Risk aversion and incentive effects,” *The American Economic Review*, 92(5), 1644–1655.
- HSU, L.-C. (2008): “Experimental evidence on tax compliance and voluntary public good provision,” *National Tax Journal*, pp. 205–223.

- (2013): “Tax auditing as a public good game: An experimental study on punishment and compliance,” *Pacific Economic Review*, 18(4), 475–501.
- JACOBSEN, C., T. R. FOSGAARD, AND D. PASCUAL-EZAMA (2017): “Why do we lie? A practical guide to the dishonesty literature,” *Journal of Economic Surveys*, 32(2), 357–387.
- JACQUEMET, N., S. LUCHINI, A. MALÉZIEUX, AND J. SHOGREN (2020): “Who’ll stop lying under oath? Empirical evidence from tax evasion games,” *European Economic Review*, 124.
- JACQUEMET, N., S. LUCHINI, A. MALÉZIEUX, AND J. F. SHOGREN (2019): “A psychometric investigation of the personality traits underlying individual tax morale,” *The B.E. Journal of Economic Analysis & Policy*, 19(3), 1935–1682.
- JOHNS, A., AND J. SLEMROD (2010): “The distribution of income tax noncompliance,” *National Tax Journal*, 63(3), 397–418.
- KAHNEMAN, D., AND A. TVERSKY (1979): “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, 47(2), 263–292.
- KASTLUNGER, B., S. G. DRESSLER, E. KIRCHLER, L. MITTONE, AND M. VORACEK (2010): “Sex differences in tax compliance: Differentiating between demographic sex, gender-role orientation, and prenatal masculinization (2D: 4D),” *Journal of Economic Psychology*, 31(4), 542–552.
- KASTLUNGER, B., E. KIRCHLER, L. MITTONE, AND J. PITTERS (2009): “Sequences of audits, tax compliance, and taxpaying strategies,” *Journal of Economic Psychology*, 30(3), 405–418.
- KASTLUNGER, B., S. MUEHLBACHER, E. KIRCHLER, AND L. MITTONE (2011): “What goes around comes around? Experimental evidence of the effect of rewards on tax compliance,” *Public Finance Review*, 39(1), 150–167.
- KIRCHLER, E. (2007): *The Economic Psychology of Tax Behaviour*. Cambridge: Cambridge University Press.
- KIRCHLER, E., S. MUEHLBACHER, E. HOELZL, AND P. WEBLEY (2009): “Effort and aspirations in tax evasion: Experimental evidence,” *Applied Psychology*, 58(3), 488–507.
- KOGLER, C., L. MITTONE, AND E. KIRCHLER (2016): “Delayed feedback on tax audits affects compliance and fairness perceptions,” *Journal of Economic Behavior & Organization*, 124, 81–87.
- LAMBERTON, C. P., J.-E. DE NEVE, AND M. I. NORTON (2018): “The power of voice in stimulating morality: Eliciting taxpayer preferences increases tax compliance,” *Journal of Consumer Psychology*, 28(2), 310–328.
- LEFEBVRE, M., P. PESTIEAU, A. RIEDL, AND M.-C. VILLEVAL (2015): “Tax evasion and social information: an experiment in Belgium, France, and the Netherlands,” *International Tax and Public Finance*, 22(3), 401–425.
- LEVITT, S. D., AND J. A. LIST (2007): “What do laboratory experiments measuring social preferences reveal about the real world?,” *The Journal of Economic Perspectives*, 21(2), 153–174.
- LIST, J. A. (2003): “Does market experience eliminate market anomalies?,” *The Quarterly Journal of Economics*, 118(1), 41–71.
- LOHSE, T., AND S. QARI (2014): “Gender differences in deception behaviour—the role of the counterpart,” *Applied Economics Letters*, 21(10), 702–705.
- LUTTMER, E. F. P., AND M. SINGHAL (2014): “Tax morale,” *The Journal of Economic Perspectives*, 28(4), 149–168.

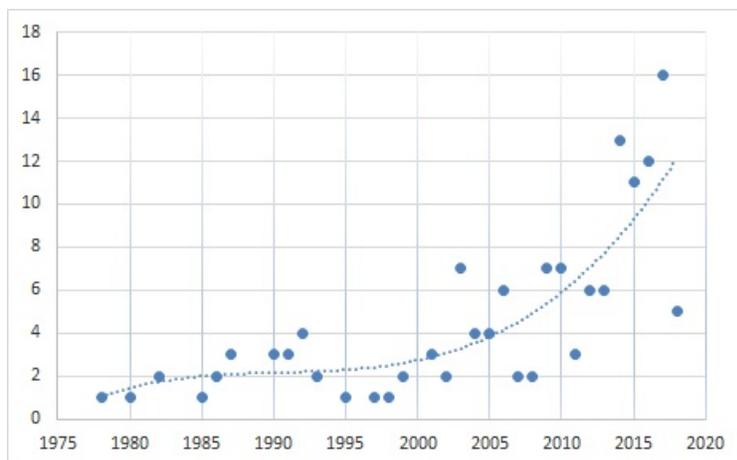
- MACIEJOVSKY, B., E. KIRCHLER, AND H. SCHWARZENBERGER (2007): "Misperception of chance and loss repair: On the dynamics of tax compliance," *Journal of Economic Psychology*, 28(6), 678–691.
- MAKI, A., M. A. COHEN, AND M. P. VANDENBERGH (2018): "Using meta-analysis in the social sciences to improve environmental policy," in *Handbook of Sustainability and Social Science Research*, pp. 27–43. Springer.
- MALÉZIEUX, A. (2016): "An investigation of the direct democracy effect of voting on tax fund destination," Working Paper.
- MALÉZIEUX, A. (2018): "A practical guide to setting up your tax evasion game," *Journal of Tax Administration*, 4(1), 107–127.
- MALIK, S., B. MIHM, AND F. TIMME (2018): "An experimental analysis of tax avoidance policies," *International Tax and Public Finance*, 25(1), 200–239.
- MASCLET, D., C. MONTMARQUETTE, AND N. VIENNOT-BRIOT (2013): "Comment réduire la fraude fiscale? une expérience sur le signalement [How to reduce tax fraud? an experiment on signaling]," Discussion paper, Centre interuniversitaire de recherche en analyse des organisations.
- MAZAR, N., O. AMIR, AND D. ARIELY (2008): "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance," *Journal of Marketing Research*, 45(6), 633–644.
- McKEE, M., C. A. SILADKE, AND C. A. VOSSLER (2018): "Behavioral dynamics of tax compliance when taxpayer assistance services are available," *International Tax and Public Finance*, 25(3), 722–756.
- MENDOZA, J. P., J. L. WIELHOUWER, AND E. KIRCHLER (2017): "The backfiring effect of auditing on tax compliance," *Journal of Economic Psychology*, 62, 284–294.
- MENKHOFF, L., AND S. SAKHA (2017): "Estimating risky behavior with multiple-item risk measures," *Journal of Economic Psychology*, 59, 59–86.
- MITTONE, L. (2006): "Dynamic behaviour in tax evasion: An experimental approach," *Journal of Socio-Economics*, 35(5), 813–835.
- MITTONE, L., F. PANEBIANCO, AND A. SANTORO (2017): "The bomb-crater effect of tax audits: Beyond the misperception of chance," *Journal of Economic Psychology*, 61, 225–243.
- MITTONE, L., AND V. SAREDI (2016): "Commitment to tax compliance: Timing effect on willingness to evade," *Journal of Economic Psychology*, 53, 99–117.
- MUEHLBACHER, S., B. HARTL, AND E. KIRCHLER (2017): "Mental accounting and tax compliance: Experimental evidence for the effect of mental segregation of tax due and revenue on compliance," *Public Finance Review*, 45(1), 118–139.
- MUEHLBACHER, S., AND E. KIRCHLER (2016): "About the external validity of laboratory experiments in tax compliance research," *Die Betriebswirtschaft*, 76(1), 7–19.
- MUEHLBACHER, S., L. MITTONE, B. KASTLUNGER, AND E. KIRCHLER (2012): "Uncertainty resolution in tax experiments: Why waiting for an audit increases compliance," *The Journal of Socio-Economics*, 41(3), 289–291.
- MUKAKA, M. M. (2012): "A guide to appropriate use of correlation coefficient in medical research," *Malawi Medical Journal*, 24(3), 69–71.

- OTTONE, S., F. PONZANO, AND G. ANDRIGHETTO (2018): "Tax compliance under different institutional settings in the EU: An experimental analysis," *Economia Politica*, 35(2), 367–402.
- PARK, C.-G., AND J. K. HYUN (2003): "Examining the determinants of tax compliance by experimental data: A case of Korea," *Journal of Policy Modeling*, 25(8), 673–684.
- PEDRONI, A., R. FREY, A. BRUHIN, G. DUTILH, R. HERTWIG, AND J. RIESKAMP (2017): "The risk elicitation puzzle," *Nature Human Behaviour*, 1(11), 803–809.
- PELIOVA, J. (2015): "Experimental investigation of factors influencing the willingness to pay taxes," in *Proceedings of the 17th International Scientific Conference Finance and Risks 2015*, vol. 1, pp. 228–232.
- RABIN, M. (2000): "Risk aversion and Expected-Utility theory: A calibration theorem," *Econometrica*, 68(5), 1281–1292.
- RECHBERGER, S., M. HARTNER, E. KIRCHLER, AND F. K. HÄMMERLE (2010): "Tax amnesties, justice perceptions, and filing behavior: A simulation study," *Law & Policy*, 32(2), 214–225.
- RILEY, R. D., P. C. LAMBERT, J. A. STAESSEN, J. WANG, F. GUEYFFIER, L. THUIS, AND F. BOUTITIE (2008): "Meta-analysis of continuous outcomes combining individual patient data and aggregate data," *Statistics in Medicine*, 27(11), 1870–1893.
- SILVERMAN, D., J. SLEMROD, AND N. ULER (2014): "Distinguishing the role of authority "in" and authority "to"," *Journal of Public Economics*, 113, 32–42.
- SLEMROD, J. (2019): "Tax compliance and enforcement," *The Journal of Economic Literature*, 57(4), 904–54.
- SNOW, A., AND R. S. WARREN (2005): "Ambiguity about audit probability, tax compliance, and taxpayer welfare," *Economic Inquiry*, 43(4), 865–871.
- ŠPALEK, J., AND Z. ŠPAČKOVÁ (2016): "Daňové úniky v laboratoři: vliv kontextu na ochotu platit daně [Tax evasion in the lab: How context affects tax compliance]," *Politická ekonomie*, 2016(5), 524–540.
- SPICER, M. W., AND L. A. BECKER (1980): "Fiscal inequity and tax evasion: An experimental approach," *National Tax Journal*, 33(2), 171–175.
- SRINIVASAN, T. (1973): "Tax evasion: A model," *Journal of Public Economics*, 2(4), 339–346.
- STEWART, L. A., M. CLARKE, M. ROVERS, R. D. RILEY, M. SIMMONDS, G. STEWART, AND J. F. TIERNEY (2015): "Preferred reporting items for a systematic review and meta-analysis of individual participant data: the PRISMA-IPD statement," *Jama*, 313(16), 1657–1665.
- TAN, F., AND A. YIM (2014): "Can strategic uncertainty help deter tax evasion? An experiment on auditing rules," *Journal of Economic Psychology*, 40, 161–174.
- TANAKA, T., C. F. CAMERER, AND Q. NGUYEN (2010): "Risk and time preferences: Linking experimental and household survey data from Vietnam," *The American Economic Review*, 100(1), 557–571.
- TORGLER, B. (2002): "Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments," *Journal of Economic Surveys*, 16(5), 657–683.
- TORGLER, B. (2003): "Beyond punishment: A tax compliance experiment with taxpayers in Costa Rica," *Revista de Análisis Económico*, 18(1), 27–56.

- (2016): “Tax compliance and data: What is available and what is needed,” *Australian Economic Review*, 49(3), 352–364.
- TORGLER, B., AND C. A. SCHALTEGGER (2005): “Tax amnesties and political participation,” *Public Finance Review*, 33(3), 403–431.
- TRIVEDI, V. U., AND J. O. CHUNG (2006): “The impact of compensation level and context on income reporting behavior in the laboratory,” *Behavioral Research in Accounting*, 18(1), 167–183.
- VON GAUDECKER, H.-M., A. VAN SOEST, AND E. WENGSTROM (2011): “Heterogeneity in risky choice behavior in a broad population,” *The American Economic Review*, 101(2), 664–694.
- VOSSLER, C. A., AND S. M. GILPATRIC (2018): “Endogenous audits, uncertainty, and taxpayer assistance services: Theory and experiments,” *Journal of Public Economics*, 165, 217–229.
- VOSSLER, C. A., AND M. MCKEE (2017): “Efficient tax reporting: The effects of taxpayer liability information services,” *Economic Inquiry*, 55(2), 920–940.
- WAHL, I., B. KASTLUNGER, AND E. KIRCHLER (2010): “Trust in authorities and power to enforce tax compliance: An empirical analysis of the “Slippery Slope Framework”,” *Law & Policy*, 32(4), 383–406.
- WAHL, I., S. MUEHLBACHER, AND E. KIRCHLER (2010): “The impact of voting on tax payments,” *Kyklos*, 63(1), 144–158.
- WAKOLBINGER, F., AND S. D. HAIGNER (2009): “Peer advice in a tax-evasion experiment,” *Economics Bulletin*, 29(3), 1653–1669.
- WARTICK, M. L., S. A. MADEO, AND C. C. VINES (1999): “Reward dominance in tax-reporting experiments: The role of context,” *Journal of the American Taxation Association*, 21(1), 20–31.
- WEBLEY, P., AND S. HALSTEAD (1986): “Tax evasion on the micro: Significant simulations or expedient experiments?,” *Journal of Interdisciplinary Economics*, 1(2), 87–100.
- YITZHAKI, S. (1974): “A note on income tax evasion: A theoretical analysis,” *Journal of Public Economics*, 3(2), 201–202.
- ZHANG, N., G. ANDRIGHETTO, S. OTTONE, F. PONZANO, AND S. STEINMO (2016): “Willing to pay? Tax compliance in Britain and Italy: An experimental analysis,” *PloS One*, 11(2).
- ZHOU, W., AND J. HEY (2018): “Context matters,” *Experimental Economics*, 21(4), 723–756.

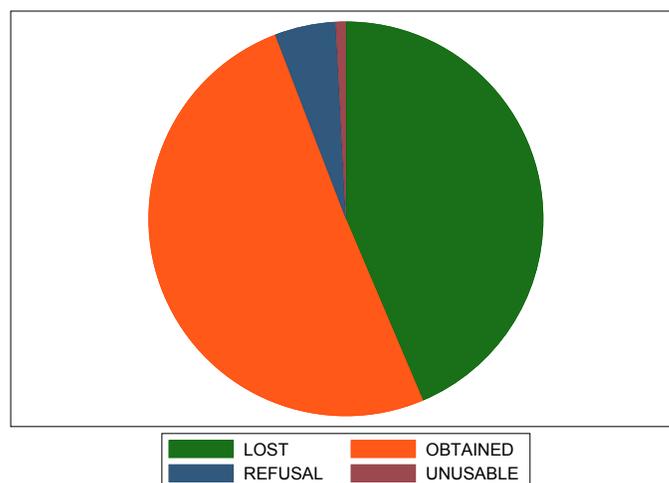
## A Appendix

Figure 2: Evolution of number of TEG per year



**Note.** This figure depicts on the Y-axis the number of TEG published or run, per year on the X-axis.

Figure 3: Status of the different articles featuring a TEG



**Note.** The whole pie represents the 137 TEG published or run until 2018. The orange part corresponds to the data obtained, the green part, to the data lost, the blue part, to the data available but that the authors did not wish to share and the red part, to one dataset that was deemed unusable.

Figure 4: First invitation to contribute

Dear [*insert name here*]

My name is Antoine Malézieux and I am a postdoc at the Tax Administration Research Centre (University of Exeter). James Alm and I are currently working on a meta-analysis of all the Tax Evasion Games published until now. To do so we want to collect the highest possible number of lab experiments where participants are asked to declare an amount of income that will be taxed afterwards. The challenge is quite huge and this is why we need your cooperation.

You have published (an) excellent article(s) [*list of articles*] in [*list of years*] in [*list of journals*]. We would very much like to add your paper(s) to our meta-analysis.

Would it be possible for you to provide us with your data in whatever format suits you? If you want to keep some parts secret, you are free to send it truncated, but of course, for the project to work, the idea is to get as much data as possible.

Best regards,

Figure 5: Reminder email

Dear [*insert name here*]

Following my email from [*insert date here*], this is just a gentle reminder. Have you had a chance to look at this yet?

Best regards,

Table 6: List of studies included in the meta-analysis

Study	# Obs.	# Clusters	Country
Abraham, Lorek, Richter, and Wrede (2017)	112	112	Germany
Adres, Vashdi, and Zalmanovitch (2016)	897	897	multiple (4)
Andrighetto, Zhang, Ottone, Ponzano, D'Attoma, and Steinmo (2016)	18,930	2106	multiple (5)
Bayer and Sutter (2009)	3,600	180	Germany
Bazart and Bonein (2014)	5,760	287	France
Beer, Kasper, and Loeprick (2016)	524	130	Austria
Blackwell and McKee (2012)	950	38	USA
Blaufus, Braune, Hundsdoerfer, and Jacob (2015)	127	127	Germany
Boylan (2010)	87	87	USA
Brockmann, Genschel, and Seelkopf (2016)	564	94	Germany
Bruner, D'Attoma, and Steinmo (2017)	18,930	2106	multiple (5)
Bruttel and Friehe (2014)	4,480	224	Germany
Buckenmaier, Dimant, and Mittone (2018)	4,020	201	Italy
Bühren and Kundt (2014)	450	150	Germany
Cadsby, Maynes, and Trivedi (2006)	305	305	Canada
Casal and Mittone (2016)	2,560	128	Italy
Casal, Kogler, Mittone, and Kirchler (2016)	492	123	Austria
Castro and Rizzo (2014)	1,800	60	Italy
Choo, Fonseca, and Myles (2015)	8,216	551	UK
Coricelli, Rusconi, and Villeval (2014)	960	32	France
Coricelli, Joffily, Montmarquette, and Villeval (2010)	1,440	96	France
D'Attoma (2018)	18,930	2106	multiple (5)
D'Attoma, Volintiru, and Malézieux (2020)	7,171	1233	multiple (5)
D'Attoma, Volintiru, and Steinmo (2017)	18,930	2106	multiple (5)
Dai (2016)	6,300	105	France
Doerrenberg and Duncan (2014a)	495	45	Germany
Doerrenberg and Duncan (2014b)	495	45	Germany
Doerrenberg (2015)	1,134	126	Germany
Doerrenberg, Duncan, and Zeppenfeld (2015)	90	90	Germany
Dulleck, Fooker, Newton, Ristl, Schaffner, and Torgler (2016)	3,420	180	Australia
Durham, Manly, and Ritsema (2014)	768	96	USA
Eisenhauer, Geide-Stevenson, and Ferro (2011)	1,096	107	USA
Etchart-Vincent and Taugourdeau (2018)	720	240	France
Fortin, Lacroix, and Villeval (2007)	670	132	France
Gërxhani and Schram (2006)	1,180	236	multiple (2)

Study	# Obs.	# Clusters	Country
Guerra and Harrington (2018)	1,620	180	multiple (2)
Guth and Sausgruber (2008)	3,375	75	Germany
Hartl, Hofmann, Gangl, Hartner-Tiefenthaler, and Kirchler (2015)	26,880	726	Austria
Heinemann and Kocher (2013)	1,600	80	Germany
Hsu (2008)	2,736	228	Taiwan
Hsu (2013)	4,204	216	Taiwan
Jacquemet, Luchini, Malézieux, and Shogren (2020)	564	216	France
Jacquemet, Luchini, Malézieux, and Shogren (2019)	113	113	France
Kastlunger, Kirchler, Mittone, and Pitters (2009)	10,800	180	Italy
Kastlunger, Dressler, Kirchler, Mittone, and Voracek (2010)	6,420	107	Italy
Kastlunger, Muehlbacher, Kirchler, and Mittone (2011)	5,160	86	Italy
Kirchler, Muehlbacher, Hoelzl, and Webley (2009)	912	304	Austria
Kogler, Mittone, and Kirchler (2016)	2,772	126	Austria
Lamberton, De Neve, and Norton (2018)	176	176	USA
Lefebvre, Pestieau, Riedl, and Villeval (2015)	7,710	257	multiple (3)
Lohse and Qari (2014)	827	261	Germany
Malézieux (2016)	500	250	France
Malik, Mihm, and Timme (2018)	133	133	Germany
McKee, Siladke, and Vossler (2018)	14,454	730	USA
Mittone (2006)	20,759	360	Italy
Mittone and Saredi (2016)	5,879	98	Italy
Mittone, Panebianco, and Santoro (2017)	4,320	144	Italy
Muehlbacher, Mittone, Kastlunger, and Kirchler (2012)	52	52	Italy
Muehlbacher, Hartl, and Kirchler (2017)	1,152	128	Italy
Ottone, Ponzano, and Andrighetto (2018)	18,930	2106	multiple (5)
Peliova (2015)	490	49	Slovakia
Rechberger, Hartner, Kirchler, and Hämmerle (2010)	2,385	238	Austria
Špalek and Špačková (2016)	2,400	120	Czech Republic
Fisar, Reggiani, and Špalek (2018)	1,100	220	Czech Republic
Silverman, Slemrod, and Uler (2014)	7,280	364	USA
Tan and Yim (2014)	3,840	128	Netherlands
Wakolbinger and Haigner (2009)	3,200	160	Austria
Wahl, Muehlbacher, and Kirchler (2010)	4,580	228	Austria
Wahl, Kastlunger, and Kirchler (2010)	4,020	201	Austria
Vossler and McKee (2017)	9,624	486	USA
Vossler and Gilpatric (2018)	17,096	714	USA
Zhang, Andrighetto, Ottone, Ponzano, and Steinmo (2016)	18,930	2106	multiple (5)

**Note.** List of publications corresponding to datasets included in this meta-analysis. One dataset can be the object of different publications.

Table 7: Public policy variables on tax compliance: comparison across main theoretical models of evasion and the present meta-analysis

	Meta-analysis		Allingham and Sandmo (1972)	Yitzhaki (1974)
	Extensive margin	Intensive margin		
<b>Audit probability</b>	Positive	Depends on the Fine	Positive	
<b>Fine size</b>	Positive	Depends on the Audit	Positive	
<b>Tax rate</b>	Negative	No impact	Ambiguous	Positive

**Note.** Comparison of public policy variables' impact on tax compliance across theoretical models of evasion and the present meta-analysis.

Table 8: Summary statistics of compliance/complier measures across countries

Country	Compliance			Complier		
	Mean	S.D.	# Obs.	Mean	S.D.	# Obs.
Albania	.91	.26	720	.89	.30	720
Australia	.56	.40	3,420	.32	.46	3,530
Austria	.67	.38	45,377	.39	.48	45,377
Belgium	–	–	–	.62	.48	3,600
Canada	.85	.32	305	.76	.42	305
Colombia	–	–	–	.74	.43	250
Czech Republic	.44	.44	3,500	.30	.45	3,500
Denmark	.52	.46	666	.43	.49	666
France	.64	.40	16,917	.46	.49	19,107
Germany	.37	.40	12,537	.28	.45	17,251
Israel	–	–	–	.60	.48	273
Italy	.63	.42	62,198	.47	.49	68,077
Netherlands	.69	.45	460	.65	.47	6,220
Romania	.74	.36	3,534	.60	.48	3,534
Slovakia	.34	.44	490	.26	.44	490
Sweden	.66	.43	3,782	.55	.49	3,782
Taiwan	.61	.38	6,940	.25	.43	6,940
UK	.68	.40	14,022	.55	.49	14,022
USA	.50	.49	59,142	.74	.37	44,662
Mean	.65	.41	219,545	.45	.49	256,801

**Note.** Summary statistics on compliance/complier measures across countries.

Table 9: Multivariate regressions of compliance decisions on public policy variables interacted with the main countries

Variable	Extensive margin		Intensive margin	
	Coef.	(St. E.)	Coef.	(St. E.)
Random audit	-.19	(.158)	.08	(.054)
Audit probability	1.59***	(.319)	.28**	(.090)
Fine size	.21***	(.047)	.04***	(.011)
Audit#Fine	.71*	(.300)	.21**	(.079)
Amnesty	-5.25***	(.067)	-.01	(.023)
Flat tax	-.21**	(.076)	-.12***	(.031)
Tax rate	-.92***	(.138)	-.08	(.053)
Austria#Audit	-.04	(2.401)	.93	(1.015)
Italy#Audit	.82	(.445)	-.04	(.171)
USA#Audit	-.74	(1.807)	.04	(.660)
Austria#Fine	.02	(.120)	-.01	(.048)
Italy#Fine	-.49***	(.108)	.05	(.038)
USA#Fine	0	(.)	0	(.)
Austria#Audit#Fine	-1.70	(1.461)	.10	(.610)
Italy#Audit#Fine	-.53	(.381)	-.21	(.125)
USA#Audit#Fine	.61	(1.836)	-1.53*	(.671)
Austria#Tax rate	0	(.)	0	(.)
Italy#Tax rate	.41*	(.208)	.01	(.081)
USA#Tax rate	.42*	(.184)	.01	(.071)
Constant	.72	(.418)	.12	(.159)
# Observations	185,425		94,999	
# Clusters	14,928		10,195	
Round FE	YES		YES	
Country FE	YES		YES	
Study FE	YES		YES	
Year FE	YES		YES	

**Note.** For the extensive margin regression, we use Probit analysis. For the intensive margin regression, we use OLS analysis, estimated on the subsample of evaders. Both models report in parentheses standard errors clustered at the individual level. Some interactions have been omitted because of collinearity. Significance levels: \* : 5% \*\* : 1% \*\*\* : 0.1%

Table 10: Multivariate regressions of compliance decisions on individual-level variables  
(income PPP variable)

Variable	Extensive margin		Intensive margin	
	M. Eff.	(St. E.)	Coef.	(St. E.)
Age	5e-5	(.001)	.0001	(.001)
Male	-.03***	(.010)	-.15***	(.021)
Student	.05	(.035)	-.11*	(.054)
Income (PPP)	-.05***	(.007)	.006	(.011)
Risk averse (HL)	.02*	(.010)	.02	(.020)
Constant	--	-	.32***	(.079)
# Observations	29,420		6,908	
# Clusters	1,549		1,075	
Round FE	YES		YES	
Country FE	NO		NO	
Study FE	YES		YES	
Year FE	YES		YES	

**Note.** For the extensive margin regression, we use Probit analysis (average marginal effects). For the intensive margin regression, we use OLS analysis, estimated on the subsample of evaders. Both models report in parentheses standard errors clustered at the individual level.

Significance levels: \* : 5% \*\* : 1% \*\*\* : 0.1%

Table 11: Multivariate regressions of compliance decisions on public policy variables  
(risk aversion)

Variable	Extensive margin		Intensive margin	
	Coef.	(St. E.)	Coef.	(St. E.)
Random audit	0	(.)	0	(.)
Audit probability	1.83***	(.394)	.39**	(.128)
Fine size	.23***	(.053)	.05***	(.013)
Audit#Fine	.82*	(.356)	.26**	(.100)
Amnesty	-4.74***	(.068)	-.01	(.023)
Flat tax	-.13	(.085)	-.12***	(.031)
Tax rate	-2.47***	(.298)	-.15	(.135)
Risk averse (HL)	.05	(.053)	.03*	(.015)
Constant	.23	(.156)	.56***	(.056)
# Observations	24,968		15,430	
# Clusters	1,732		1,386	
Round FE	YES		YES	
Country FE	YES		YES	
Study FE	YES		YES	
Year FE	YES		NO	

**Note.** For the extensive margin regression, we use Probit analysis. For the intensive margin regression, we use OLS analysis, estimated on the subsample of evaders. Both models report in parentheses standard errors clustered at the individual level. Random audit has been omitted because of collinearity problem.  
Significance levels: \* : 5% \*\* : 1% \*\*\* : 0.1%

Table 12: Summary statistics on risk attitude measures

Variable	Value	SD	Min	Max	# Obs.
Risk: Holt & Laury	5.82	1.94	0	10	49,774
Risk averse (HL)	.58	.49	–	–	49,774
Risk: Subjective question (1)	5.95	2.29	0	10	24,640
Risk averse (SQ1)	.27	.44	–	–	24,640
Risk: Subjective question (3)	5.14	1.81	1	8	7,710
Risk averse (SQ3)	.40	.49	–	–	7,710
Risk: Bomb Risk Elicitation Task	46.19	19.05	7	99	5,879
Risk averse (BRET)	.44	.49	–	–	5,879
Risk: 20 Lotteries	7.54	4.66	0	19	5,760
Risk averse (20L)	.58	.49	–	–	5,080
Risk: 1 Lottery	1.30	.65	1	3	1,719
Risk averse (1L)	.79	.40	–	–	1,719
Risk: Portfolio Choice	9.37	3.57	2	15	1,440
Risk averse (PC)	.27	.44	–	–	1,440
Risk: Subjective question (2)	4.48	1.83	1	8	492
Risk averse (SQ2)	.48	.50	–	–	492
Risk: Dospert-Financial	17.04	6.91	6	42	450
Risk averse (Dospert)	.84	.36	–	–	450

**Note.** Summary statistics on the different risk attitude measures found in the literature.

Table 13: Multivariate regressions of compliance on public policy variables and experimental design variables

Variable	Extensive margin		Intensive margin	
	Coef.	(St. E.)	Coef.	(St. E.)
Random audit	.44*	(.190)	-.06	(.051)
Audit probability	1.03	(.577)	-.22	(.123)
Fine size	.23***	(.059)	.07***	(.011)
Audit#Fine	.82*	(.365)	.49***	(.081)
Amnesty	-4.86***	(.372)	-.01	(.022)
Flat tax	-.24***	(.077)	-.12***	(.030)
Tax rate	-.19	(.160)	-.007	(.052)
Loaded framing	-.002	(.157)	-.11*	(.049)
Directive way	5.74***	(.474)	.47***	(.101)
Task tax	17.52***	(.901)	.20	(.108)
Earned	8.38***	(.456)	.90***	(.117)
Redistribution	.61***	(.085)	.08***	(.023)
Public Good	.68***	(.075)	.07***	(.021)
Income	-.14***	(.010)	-.007	(.006)
Constant	-4.62***	(.514)	.40***	(.112)
# Observations	66,834		36,456	
# Clusters	5,679		4,451	
Round FE	YES		YES	
Country FE	YES		YES	
Study FE	YES		YES	
Year FE	YES		YES	

**Note.** For the extensive margin analysis, we use Probit estimation. For the intensive margin analysis, we use OLS estimation, estimated on the subsample of evaders. Both models report in parentheses standard errors clustered at the individual level. Significance levels: \* : 5% \*\* : 1% \*\*\* : 0.1%

Table 14: Multinomial regressions of behavioral types on public policy variables, experimental design variables and individual-level variables

Variable	M. Logit (1)		M. Logit (2)		M. Logit (3)	
	Coef.	(St. E.)	Coef.	(St. E.)	Coef.	(St. E.)
<b>Full evaders</b>						
Random audit	3.69***	(1.035)				
Audit probability	-4.06***	(.562)				
Fine size	-.26***	(.067)				
Audit #Fine	-1.48***	(.440)				
Amnesty	.54	(.379)				
Flat tax	1.88*	(.736)				
Tax rate	-2.20***	(.496)				
Loaded framing			-3.54***	(.442)		
Directive way			.21	(.255)		
Tax task			-.17	(.313)		
Earned			1.63***	(.177)		
Redistribution			-.87***	(.186)		
RL public good			-.74***	(.194)		
Age					.01	(.069)
Male					1.29*	(.551)
Student					3.04	(2.618)
Income					.09	(.132)
Risk averse (HL)					-.28	(.410)
Constant	-9.16***	(1.293)	-4.01***	(.427)	-7.55	(4.339)
<b>Inconsistent</b>	0	(.)	0	(.)	0	(.)
<b>Full compliers</b>						
Random audit	.03	(.366)				
Audit probability	-.37	(.377)				
Fine size	-.32***	(.046)				
Audit #Fine	-1.29***	(.312)				
Amnesty	.25	(.184)				
Flat tax	.61	(.336)				
Tax rate	-1.67***	(.392)				
Loaded framing			1.13***	(.249)		
Directive way			.95***	(.169)		
Tax task			.42*	(.214)		
Earned			.72***	(.180)		
Redistribution			1.26***	(.154)		
RL public good			.44*	(.182)		
Age					.008	(.008)
Male					-.28*	(.122)
Student					-.40	(.322)
Income					-.10	(.070)
Risk averse (HL)					.20	(.120)
Constant	-1.26*	(.524)	-3.28***	(.346)	-2.41***	(.664)
# Observations	163,172		147,660		29,420	
# Clusters	11,220		8,495		1,549	
Round FE	YES		YES		YES	
Year FE	YES		YES		YES	

**Note.** We use multinomial logistic regressions where the dependent variable is the behavioral types of participants (0= full evader, 1= people who varies between both and 2=full complier). The base are the inconsistent participants (which are omitted). The models report in parentheses standard errors clustered at the individual level. Significance levels: \* : 5% \*\* : 1% \*\*\* : 0.1%

Table 15: Multivariate regressions of compliance decisions on public policy variables, experimental design variables and individual-level variables (alternative models)

Variable	Extensive margin			Intensive margin		
	(1)	(2)	(3)	(4)	(5)	(6)
	Coef. (St. E.)	M. Eff. (St. E.)	M. Eff. (St. E.)	Coef. (St. E.)	M. Eff. (St. E.)	M. Eff. (St. E.)
Random audit	-.31 (.260)			.43 (.253)		
Audit probability	2.53*** (.246)			-1.75*** (.199)		
Fine size	.18** (.066)			-.02 (.048)		
Audit#Fine	.96*** (.229)			-1.55*** (.200)		
Amnesty	-16.18*** (.136)			-.32 (.195)		
Flat tax	-.36** (.123)			-.53*** (.134)		
Tax rate	-1.16*** (.163)			.07 (.193)		
Loaded framing		.03 (.047)			-.06 (.036)	
Directive way		.42*** (.090)			.09 (.081)	
Tax task		-.10 (.097)			.09 (.083)	
Earned income		.02 (.066)			-.08 (.060)	
Redistribution		.14*** (.023)			.08*** (.016)	
RL public good		.17*** (.028)			.08*** (.023)	
Age			-.0001 (.001)			-.0002 (.001)
Male			-.03*** (.011)			-.15*** (.020)
Student			.05 (.035)			-.11* (.051)
Income			-.05*** (.009)			.006 (.012)
Risk averse (HL)			.02* (.010)			.02 (.020)
Constant	-2.17*** (.362)	—	—	-1.53*** (.353)	—	—
# Observations	185,425	155,997	29,420	94,999	84,550	6,908
# Clusters	12,112	9,700	1,549	8,878	6,839	1,075
Round FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES	YES
Study FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

**Note.** For the extensive margin analysis, we use Logit estimation (average marginal effect in Columns 2 and 3). For the intensive margin analysis, we use GLM estimation (binomial family distribution, average marginal effect in Columns 5 and 6), estimated on the subsample of evaders. Both models report in parentheses standard errors clustered at the individual level. Significance levels: \* : 5% \*\* : 1% \*\*\* : 0.1%

Table 16: Multivariate regressions of compliance decisions on public policy variables, experimental design variables and individual-level variables  
(average effect)

Variable	Tobit						OLS					
	(1)		(2)		(3)		(4)		(5)		(6)	
	Coef.	(St. E.)	M. Eff.	(St. E.)	M. Eff.	(St. E.)	Coef.	(St. E.)	Coef.	(St. E.)	Coef.	(St. E.)
Random audit	-.07	(.138)					-.007	(.050)				
Audit probability	.49***	(.090)					.001	(.034)				
Fine size	.03	(.028)					.006	(.010)				
Audit#Fine	-.32***	(.086)					-.23***	(.033)				
Amnesty	-.78***	(.099)					-.31***	(.028)				
Flat tax	-.29***	(.059)					-.12***	(.024)				
Tax rate	-.58***	(.084)					-.17***	(.031)				
Loaded framing			.07	(.088)					-.01	(.037)		
Directive way			.90***	(.164)					.30***	(.075)		
Tax task			-.10	(.149)					-.001	(.086)		
Earned income			.03	(.099)					-.02	(.064)		
Redistribution			.29***	(.046)					.14***	(.018)		
RL public good			.35***	(.057)					.17***	(.024)		
Age					-.3e-6	(.004)					.0006	(.0009)
Male					-.14***	(.042)					-.05***	(.008)
Student					.20	(.154)					-.05	(.041)
Income					-.26***	(.032)					-.03***	(.008)
Risk averse (HL)					.10*	(.042)					.01*	(.008)
Constant	-.034	(.187)	—	—	—	—	.32***	(.068)	.50***	(.085)	.65***	(.064)
# Observations	163,123		155,997		29,420		163,123		147,611		29,420	
# Clusters	11,220		9,700		1,549		11,220		8,495		1,549	
Round FE	YES											
Country FE	YES											
Study FE	YES											
Year FE	YES											

**Note.** On the left hand side, we use Tobit estimation (average marginal effect in Columns 2 and 3, Left-censored = 0, Right-censored = 1). On the right hand side, we use OLS estimation. Both models report in parentheses standard errors clustered at the individual level.  
Significance levels: \* : 5% \*\* : 1% \*\*\* : 0.1%

Table 17: Multivariate regressions of compliance decisions on public policy variables, experimental design variables and individual-level variables (cluster study level)

Variable	Extensive margin			Intensive margin		
	(1)	(2)	(3)	(4)	(5)	(6)
	Coef. (St. E.)	M. Eff. (St. E.)	M. Eff. (St. E.)	Coef. (St. E.)	Coef. (St. E.)	Coef. (St. E.)
Random audit	-55** (.212)			.11 (.084)		
Audit probability	1.02* (.483)			-.29 (.277)		
Fine size	.07* (.036)			.04** (.015)		
Audit#Fine	.46 (.354)			.05 (.149)		
Amnesty	.11 (.728)			.16 (.191)		
Flat tax	-58*** (.149)			-.33*** (.067)		
Tax rate	-.75 (.476)			.22 (.116)		
Loaded framing		.27** (.104)			.02 (.047)	
Directive way		.02 (.049)			-.08*** (.018)	
Tax task		-.11 (.079)			.05 (.033)	
Earned income		-.06 (.064)			-.01 (.035)	
Redistribution		.07* (.032)			.04*** (.013)	
RL public good		.04 (.032)			.004 (.030)	
Age			6e-5 (.001)			.0001 (.002)
Male			-.03 (.020)			-.15* (.034)
Student			.05 (.093)			-.12* (.041)
Income			-.06 (.041)			.008 (.013)
Risk averse (HL)			.02** (.008)			.02 (.023)
Constant	-.24 (.319)	—	—	.34* (.153)	.21*** (.048)	.43* (.134)
# Observations	185,425	155,997	29,420	94,999	84,550	6,908
# Clusters	51	33	5	48	30	5
Round FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

**Note.** For the extensive margin analysis, we use Probit estimation average marginal effect for Columns 2 and 3). For the intensive margin analysis, we use OLS estimation, estimated on the subsample of evaders. Both models report in parentheses standard errors clustered at the study level.  
Significance levels: \* : 5% \*\* : 1% \*\*\* : 0.1%