CESifo WORKING PAPERS

10615
2023

August 2023

The Bright Side of Tax Evasion

Wladislaw Mill, Cornelius Schneider

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10615*

The Bright Side of Tax Evasion

Abstract

This paper investigates whether tax evasion can be beneficial for an optimal income tax schedule. Past theoretical discussions have presented mixed outcomes as to whether allowing taxpayers to opt into uncertainty could indeed enhance overall tax revenues. In this study, we conducted an original real effort experiment in an online labor market with almost 1,000 participants to test this hypothesis empirically. Our findings show significant positive labor supply responses to the opportunity to evade (increased labor supply by 37%). More importantly, the expected tax revenue significantly and substantially increased by up to more than 50%. As an example, our data suggests that a 40% tax rate with complete enforcement could be replaced with a 28% tax rate with the option of tax evasion, without any loss in tax revenue. Strikingly, this effect persists when comparing effective tax rates: Lowering effective tax rates through probabilistic enforcement (the opportunity to evade) is more efficient than simply lowering statutory tax rates. Our findings suggest that the opportunity for tax evasion can increase tax revenues beyond what a corresponding decrease in nominal rates would achieve. For welfare analyses, this highlights the importance of not only considering the elasticity of *taxable* income (ETI) but *total* earned income elasticities.

JEL-Codes: H210, H240, H260, J220, C910.

Keywords: tax evasion, tax revenues, labor supply, optimal taxation, experiment.

Wladislaw Mill
Department of Economics
University of Mannheim
L7 3-5
Germany – 68161 Mannheim
mill@uni-mannheim.de

Cornelius Schneider
Department of Economics
University of Mannheim
L7 3-5
Germany – 68161 Mannheim
schneider@uni-mannheim.de

August 14, 2023

1 Introduction

Ample literature reports how tax evasion remains a pervasive problem that has not been effectively controlled or mitigated (Bustos et al., 2022). This issue is becoming even more pronounced, with the tax gap growing by about 12% according to the latest IRS estimates.[1] It appears that an ongoing race between tax enforcement and tax planning is underway, with tax planning currently in the lead. This scenario brings problems for everyone - it's bad news for taxpayers and a big challenge for tax authorities. Previous literature typically classified the negative responses to taxation into two categories: labor supply responses and evasion/avoidance decisions. The majority of these studies have indicated that particularly the latter is relevant for assessing revenue implications (Saez et al., 2012; Gruber and Saez, 2002). Consequently, the dominant recommendation has not been to adjust tax rates, as Mirrlees (1971) suggests, but to minimize evasion opportunities. However, adjusting tax rates might be politically challenging, and eliminating evasion has proven to be difficult. A truly innovative approach to the dilemma might provide the *interplay* of the two response margins - labor supply and tax evasion - which is surprisingly understudied: if evasion opportunities imply positive labor supply responses, the strict elimination of those might not be desirable. This idea was first formalized by Weiss (1976): The government may benefit from strategically permitting a certain degree of tax evasion, as it can lead to increased tax revenue and enhance overall welfare. A subsequent theoretical debate recognized that the possibility could theoretically exist in some form, but deemed the underlying assumptions to be empirically highly improbable (Sandmo, 1981; Hellwig, 2007). Yet, whether such a counterintuitive effect exists is essentially an empirical question that has remained unanswered to date.

This project aims to fill this gap. The inconclusive, exclusively theoretical discussion on the desirability of evasion opportunities calls for an empirical assessment. The question this paper aims to answer is: can the opportunity to evade taxes increase overall tax revenue? More specifically, does a tax system that allows for some degree of evasion via probabilistic enforcement generate more tax revenue than a system with perfect enforcement?

For this, we ran an original real effort experiment in a real online labor market with nearly 1,000 participants. Importantly, in order to reveal honest labor supply responses, our participants individually decide upon the time they want to spend working on our real effort task for which they are paid on a piece-rate basis. The task participants are required to do is to solve the number of zeros in a given table. To elicit the labor supply each of our subjects is asked to indicate her willingness to work for eight different payment scenarios: Within each flat-tax rate of 20%, 40%, 60% and 80%, they are faced with either a "low wage" or a "high wage". After stating the number of tables they are willing to work

[1] (see: https://www.irs.gov/newsroom/irs-updates-tax-gap-estimates-new-data-points-the-way-toward-enhancing-taxpayer-service-compliance-efforts (13.07.23))

on in each of these eight payment scenarios, respondents are required to actually work on only one randomly picked scenario and are paid out accordingly. While this setting is completely identical in our treatment and control group, the treatment group is given the opportunity to evade the tax: They decide upon their labor supply while simultaneously being able to invest (part of) their income into a lottery. If participants win, they avoid paying any tax on the invested amount; if they lose, they have to pay the tax on the invested amount plus an additional fine of 20% on the money invested. The subjects in the control group are only able to decide upon their labor supply, without any opportunity to avoid the tax. Collecting eight different data points per participant on labor supply allows us to elicit individual reservation wages and labor supply elasticities while avoiding undesired effects of tediousness between the different scenarios.

This online-lab experimental design also offers a couple of further advantages: First, the labor market characteristics allow us to elicit actual willingness to work: each participant is free to decide on the amount of time spent on the experiment herself. Respondents' participation in the experiment is only granted as long as the payment scheme is perceived as profitable. This constitutes a methodological advantage to related studies that implemented real effort tasks only in true lab-settings with fixed time frames (e.g. Doerrenberg and Duncan, 2014). Next, such a gamified and artificial experiment allows us to cleanly identify the core-mechanism suggested by the theory. Foremost, the two response margins of labor supply and evasion decisions can be distinctively disentangled, a common challenge in the ETI literature mostly relying on field data. It further deliberately abstracts from many real-life aspects, canceling out confounding factors like framing, lying aversion, moral costs or social preferences (as discussed in e.g. Slemrod, 2007; Kirchler et al., 2008).

Our empirical findings are as follows: First, we find in general strong labor supply responses to changes in the net-of-tax rate. Under the most profitable condition (high wage and low tax) only 7% of all participants decided not to work, whereas under the worst condition (low wage and high tax) already 34% of all participants decided not to work. The average labor supply – aggregated over all tax levels and the two wage levels – in the Evasion-treatment increased substantially, on average by 37%, compared to the situation without the opportunity for evasion. Second, we find that a considerable amount is evaded - on average almost 40% of the income. Finally, and most important, we find that the opportunity to evade significantly and substantially increases the expected tax revenue, by more than 50% in the highest tax scenario. Whereas we document a classical Laffer curve with peak at a 60% tax in our control treatment, the substitution effect between work and leisure seems to be offset and tax revenues keep increasing with increasing tax rates (highest tax revenues at 80%) in our Evasion-treatment. This effect is strongest for the low-wage scenario. Strikingly, this effect still prevails when comparing effective tax rates: Lowering effective tax rates through the opportunity to evade is more efficient than simply lowering statutory tax rates.

As our experiment design abstracts from numerous real-world factors, concerns about external validity may arise. The elasticities of labor supply observed in our online sample are, however, consistent with those found in observational studies in real-world scenarios.[2] One of our pivotal assumptions is the non-rigidity and complete flexibility of labor markets, positing that individuals can adjust their labor supply continuously and without restrictions. This flexibility, aligns with the observational data reported for a substantial subset of the population. Specifically, our findings are pertinent to self-employed individuals and small businesses, who have the capability to modify their labor supply, thereby reinforcing the relevance of our results. Recent official tax gap estimates by the Internal Revenue Service (IRS) (2020) reflect the significance of this group. They quantify the average annual gross tax gap[3] between 2011 and 2013 at $441 billion, with a non-compliance rate of 16%. To put this into perspective: this is as much as half of the US deficit in 2012, or, the federal spending on health and education taken together (Office of Management and Budget, 2013). This gap is primarily driven by nonfiling and underreporting in individual incomes and self-employed individuals, accounting for 80% of the total tax gap (Internal Revenue Service (IRS), 2019). Recent empirical research supports this picture, with workers exploiting information asymmetries with substantial tax evasion when the true tax base lacks observability (Kleven et al., 2011; Slemrod, 2007) (i.e. self-reported incomes and sectors that rely on cash). This behavior is observable throughout the whole income distribution, from plumbers to lawyers (Alstadsæter et al., 2019). These empirical observations present a compelling basis for the potential labor supply-enhancing effects of tax evasion opportunities to be of considerable significance and relevance.

Although the issue of income tax evasion and avoidance having received extensive theoretical and empirical attention over the past few decades, previous empirical literature did not explore our proposed mechanism. Given the conventional view of tax evasion as unjust and inefficient, which leads to diminished tax revenues, previous empirical research has primarily emphasized the importance of enhancing tax compliance as a strategy to raise revenue (Slemrod, 2019).[4] Assuming that individuals behave rationally and selfishly (Becker, 1968), research has consistently shown that increased fines and audit probabilities reduce tax evasion, building upon the framework developed by Allingham and Sandmo (1972) (Blackwell, 2010; Mascagni, 2018). This finding has been replicated across lab experiments, field experiments, and administrative data (Torgler, 2002; Pomeranz, 2015; Alstadsæter et al., 2022). Other studies have explored the effects of intrinsic motivation (Wahl et al., 2010), social norms and morals (Fellner et al., 2013), and (mis)perceived audit probabilities (Bott et al., 2019).

The seminal theoretical work by Weiss (1976) deems this perspective as too short-sighted. He shows how an optimal income tax might include incentives to cheat in order

[2]The elasticities of labor supply range between 0.19 and 1.3 in our sample, very close to those found in observational data by e.g. Heim (2010) or Chetty (2009).

[3]Estimated total true tax liability minus taxes paid voluntarily and timely.

[4]With the only reason why tax evasion might be acceptable being disproportionate costs of enforcement (see for example Keen and Slemrod, 2017).

to partially offset the undesirable distortions of taxes on labor supply. For risk-neutral individuals, allowing tax evasion translates into an effective tax reduction down to the expected tax payment. However, under specific assumptions about individual risk-aversion (i.e. prudence; Kimball, 1990), evasion opportunities (or, put differently, probabilistic enforcement) might lead to a disproportionately increased labor supply, resulting in higher tax revenues than achieved without such evasion opportunities. In short, the effect considers how taxpayers respond to uncertainty: behavior may differ significantly between risk-averse and risk-neutral taxpayers, potentially leading to an increase in revenue. Most importantly, Weiss (1976) argues that allowing for tax evasion can have a welfare-increasing effect *beyond* what would be achieved by simply reducing the statutory tax rates equivalently. That is, the opportunity to cheat can reduce the excess burden of a tax system (i.e. by increasing labor supply), while tax revenues are being held constant or even increasing - representing a real Pareto improvement. More generally speaking, the opportunity to evade offers a mechanism for tax-sensitive individuals to self-select into a more "personalized" tax scheme. Meanwhile, the social planner maximizes tax revenues by optimizing these individuals' labor supply, exploiting their risk preferences.

Remarkably, the essential components for Weiss's mechanism to function were previously identified in a distinct, albeit related, paper by Block and Heineke (1973). Adopting a more general approach, they focused on labor supply under increased income uncertainty. Their study reveals that, depending on individual risk aversion, people may employ hedging and self-insurance strategies to counteract fluctuations in labor income by increasing their labor supply. Consequently, they conclude that policies aimed at reducing uncertainty could lead to disincentive effects, while a carefully designed increase in income uncertainty might serve as a policy instrument to stimulate labor supply. Subsequent works discuss this mechanism under the concept of "randomized taxation" with the "desirability of cheating" as a special case. Stiglitz (1982), Brito et al. (1995) and Hellwig (2007) discuss the necessary assumptions for tax evasion opportunities to be optimal. In essence, these are twofold. First, labor supply needs to respond to taxes. The more a decrease in taxes increases labor supply, the greater the revenue from decreasing effective tax rates through the opportunity to evade. Second, risk needs to increase labor supply. Specifically, the induced uncertainty only increases tax revenues, if individual' preferences exhibit a property of absolute risk aversion decreasing in income.[5] In essence, by lowering effective tax rates through the opportunity to evade, one can exploit taxpayers' prudence. The presence of background risk results in an increase in labor supply that outweighs the loss in revenue from tax evasion. Whether this condition holds is essentially an empirical

[5]The FOC can be represented as $E(u_c * w - u_l)$, where E denotes the expected value. For uncertainty regarding disposable income to increase this expected value beyond zero (in other words, to encourage more work), the expression within the expectation operator should display convexity in relation to disposable income. When consumption (c) and leisure are additive, the requirement for this convexity is that the third derivative of the utility function with respect to consumption u_{ccc} is greater than zero, which signifies prudence. Prudence is a necessary condition for decreasing absolute risk aversion (DARA).

question. While previous literature has been strictly theoretical in discussing this mechanism and casting doubts on the likelihood of its realization in real-world circumstances (Sandmo, 1981; Yitzhaki, 1987; Schroyen, 1997; Hellwig, 2007), its empirical relevance has never been directly tested. Our study adds a significant contribution to this body of research, as we provide empirical evidence supporting this mechanism.

Building on these theoretical foundations, we present the first empirical test of the direct link between the opportunity to evade taxes and its impact on overall tax revenues. To the best of our knowledge, this is the first study to do so. Although the prominent literature on the elasticity of taxable income (ETI) specifies labor supply and evasion/avoidance as the two relevant response margins, it relies largely on the strong assumption that these need not the be differentiated to determine the welfare costs of an income tax (Feldstein, 1999). Against this background, a large body of empirical literature estimated ETIs (mostly on self-employed or high-income taxpayers, due to their higher number of opportunities adjusting their income) (Neisser, 2021), but without distinguishing differential revenue implications of these two margins. As in: they do not take into account the transfer of non-taxed income (such as fines) or, crucially for our study, the possible interplay between evasion and labor supply responses. Further, this literature faces considerable empirical challenges "untangling tax-related from non-tax-related changes in reported incomes" with administrative data (Saez et al., 2012, p. 33).

In terms of the interaction between labor supply responses and evasion opportunities, the work by Doerrenberg and Duncan (2014) is more closely related to our paper. They investigate the effect that tax evasion opportunities have on labor supply in laboratory experiments. They find that labor supply indeed responds to opportunity to evade. However, their effect (both in size and sign) strongly hinges on the specific tax rate evolution over time and remains fairly small. Our study differs from the latter paper in a fundamental way: they assess labor supply responses and do not elaborate on overall tax revenues. Even though (positive) labor supply responses represent an important aspect to the proposed theoretical mechanism, this mechanism also comprises a more comprehensive welfare perspective including i.e. the role of paid penalties and risk aversion.

In terms of the interaction between tax revenues and evasion opportunities, a couple of papers in Development Economics analyzed how to raise tax revenues when tax evasion is inevitable; i.e. facing limited state capacities (Emran and Stiglitz, 2005; Best et al., 2015). A recent paper by Bergeron et al. (2021) found property tax rates in the D.R. Congo to be too high with little compliance on the extensive margin, due to liquidity constraints. They propose sequentially increasing enforcement and tax rates to optimize the revenue-maximizing tax rate. However, they remain unclear about the absolute optimal level of enforcement and, even more important for our paper, the property tax does not allow for responses on the intensive margin (i.e. adjusting labor supply).

Our empirical support of the mechanism suggested by Weiss (1976) speaks fundamentally to a variety of research areas. We view our main contribution as being threefold.

First, and most substantial, our findings have fundamental implications for the welfare analysis of tax evasion. The following studies already highlighted potential positive effects of tax evasion, however, these are limited to rather static considerations. The standard literature on optimal administrative tax enforcement usually equates the marginal costs of increased tax enforcement (e.g an additional tax official) to the marginal revenue gains (for a comprehensive overview see: Slemrod, 2019). Nevertheless, these studies consider the tax basis as a datum and ignore its endogeneity. Two studies go beyond this mechanism: Keen and Slemrod (2017) argue to also factor in real labor supply responses of increased enforcement. However, they only consider the negative labor supply responses of increased enforcement to the extent an equivalent explicit tax rate increase would have. Slemrod and Traxler (2010) connect to the literature of Weiss (1976) and Hellwig (2007) acknowledging the possibility of labor supply responses beyond what a similar increase in tax rates would imply. As a result, their theoretical work already anticipates the possibility of the desirability of allowing for some extent of non-observability of the tax base, *even if enforcement would be costless*. Yet, this theoretical eventuality lacks empirical support. Our empirical findings strongly support this notion: tax evasion opportunities are still desirable, even if perfect enforcement would be costless. Moreover, in a more fundamental sense, our findings also contribute to the closely connected literature on the deadweight loss of income taxation. According to the seminal work of Feldstein (1999), the efficiency costs of decreased labor supply and tax evasion are equivalent. However, Chetty (2009) objects to this, arguing that the deadweight loss arising from evasion and avoidance is significantly smaller when taking into account *sheltering costs*, such as charitable giving, setting up trusts, and hiring tax consultants, compared to the deadweight loss caused by considering merely decreased reported, taxable income. Chetty suggests that both the elasticity of taxable income and the elasticity of *total* earned income should be considered when assessing the welfare implications of a tax. In conclusion, the taxable income elasticity is not sufficient to calculate deadweight loss and a comprehensive understanding of the implications of evasion and avoidance is crucial for a complete assessment. Our results directly connect to this rationale of Chetty (2009), adding increased labor supply to his perspective on *sheltering costs*. If the increase in labor supply and compliance outweighs the loss in revenue due to tax evasion, the government can collect more tax revenue with a lower effective tax rate. This would imply that the excess burden of taxation is not as large as Feldstein argues, as the presence of evasion opportunities allows for a more efficient tax collection mechanism by exploiting taxpayers' risk aversion. Therefore, the emphasis on the concept of *taxable* income elasticity may overemphasize the negative implications of tax evasion. Along similar lines, conventional measures might overstate tax gaps due to the tax-base-reducing behavioral effects triggered by perfect enforcement. In its consequence, conventional tax gap measures are potentially based on misleading counterfactuals.[6]

[6] Gemmell and Hasseldine (2014) already discuss such an effect - but not beyond what would be

Second, our results directly speak to the theoretical debate on the desirability of evasion opportunities (i.e. "randomized taxation"). Specifically, Hellwig (2007) posits that a stochastic tax scheme is only desirable assuming a specific type of risk aversion, which he considers to be empirically improbable. On the contrary, our experiment strongly suggests the empirical existence of this mechanism, yielding efficiency gains by inducing uncertainty. The exhibited positive labor supply responses are so pronounced that overall tax revenues actually increase. Put in terms of *certainty equivalents*: individuals are even willing to pay a price to enjoy the uncertainty induced by the opportunity to evade; despite the prevailing intuition that this mechanism might be considered a curiosity. Thus, our empirical findings also support mechanisms that have not yet been sufficiently reflected in the field of tax evasion. Particularly, our study observes a significant proportion of participants opting for tax evasion, coupled with a strategy of increasing their labor supply. This dual behavior can be interpreted as self-selection into evasion, whilst simultaneously hedging against the potential risk of penalties. Related fields provide clear indicators that can effectively explain this mechanism: This issue can be considered a subquestion within the broader literature on optimal decision-making in response to risk (e.g. "precautionary saving"; Kimball, 1990), including precautionary saving, prudence, and higher-order risk aversion (e.g. Deck and Schlesinger, 2010; Noussair et al., 2014; Ebert and Wiesen, 2014). In our context, Block and Heineke (1973) propose such a mechanism under the term "uncertainty effect". They posit that individuals facing income uncertainty - akin to those engaging in tax evasion - might bolster their labor supply as a hedge, increasing their expected income to mitigate the risk of potential penalties. This self-insurance behavior, in accordance with prudence, becomes increasingly apparent as income variability intensifies. For the self-selection into evasion, the literature suggests a potential interplay of several factors. Second-order risk aversion (Segal and Spivak, 1990) could lead individuals to be risk-neutral towards small perceived risks like evasion. Expectations-based reference points (Kőszegi and Rabin, 2006) may shift individuals' perceptions of risk and reward. The heterogeneity of gain-loss attitudes (Campos-Mercade et al., 2022) underscores that a significant subset of individuals might even exhibit risk-loving tendencies. Importantly, allowing for tax evasion does not entail inducing general uncertainty but rather providing an instrument for individuals to self-select into tax evasion if they desire. Policymakers could exploit this phenomenon to achieve a Pareto improvement, effectively catering to the preferences of different taxpayers.

Finally, our study seeks to provide valuable insights for policymakers to optimize revenue collection strategies while minimizing negative economic impacts. The examined mechanism raises the question of how this affects the behavior of workers who can adjust their labor supply flexibly and continuously, such as vendors, cleaners, plumbers, and hairdressers. Our findings imply that governments might be well-advised to refrain

achieved by the enforcement-equivalent tax increase.

from enforcing absolute tax compliance among its citizens. This is particularly relevant in light of the increasing prevalence of cashless economies.[7] Notwithstanding, there are already cases in which governments tolerate a certain degree of evasion and avoidance behavior: fiscal competition. If statutory tax rates are set by a federal government, while tax collection is implemented by the local authority, these authorities can determine their actual tax rate through their enforcement policy (e.g., through the number of tax officials they employ or the deductions they grant).[8] Previous works associate this behavior with the conventional notion of efficient tax administration. Our findings add another rationale to decrease the effective tax rate: tax base increasing effects and thus potentially higher revenues. Furthermore, our results imply a differential treatment of different income levels. This finding yields a policy recommendation for the most prevalent case: Due to limited resources, governments are just not capable of fully observing the tax base. Since the positive effect on labor supply is more pronounced for low-wage earners, governments should rather focus their resources on the enforcement of high-wage earners (For an insightful theoretical discussion on the welfare-enhancing focus of limited governmental resources in tax enforcement, see Lederman and Sichelman, 2013). Given this is mostly already common practice, our paper provides another economic perspective on this implicit rationale.

The remainder of the paper proceeds as follows: Section 2 provides a brief model with predictions, sections 3 and 4 will give a detailed description of our experimental design and the data. In Section 5 the results are presented followed by a brief discussion and concluding remarks in Section 6.

2 Motivating Framework

In this section, we present the theoretical model of Weiss (1976) who extends the seminal framework of Allingham and Sandmo (1972) by integrating the decision on reporting income with the decision on labor supply. This framework serves to illustrate how the determinants of labor supply and tax evasion affect overall tax revenues. The subsequent experimental design aims to match this framework and operationalizes these determinants. Since our paper is not a direct test of this theory, we introduce the intuition rather briefly.

[7] The advent of digital currencies has brought about a paradigm shift in the way financial transactions are carried out. Unlike their decentralized and anonymous counterparts, these currencies provide increased observability of the tax base, thereby enabling tax enforcement to become almost costless. This has been exemplified by the recent introduction of the E-Krona in Sweden and the E-Yuan in China, which have pushed these countries closer towards a fully cashless economy. In fact, statistics from 2018 reveal that cash accounted for only 10% of transactions in Sweden, indicating a strong move towards a digital economy (see: `https://www.nytimes.com/2018/11/21/business/sweden-cashless-society.html` (13.07.23)). A trend that has been massively accelerated during the Covid-19 pandemic.

[8] See Stö and Traxler (2005) for conceptual considerations; Cremer and Gahvari (2000), Bönke et al. (2017), Baretti et al. (2002) for empirical references.

For formal derivations please see Weiss (1976).

2.1 Model and Hypotheses

Individuals decide on their labor supply l and the share of underreported income $a \in [0, 1]$, given a certain wage rate ω, which we normalize to 1 for simplicity. The resulting pre-tax income $c = l$ is taxed by the government at a rate $t \in [0, 1]$, leading to an after-tax income of $l(1-t)$. However, individuals also have the opportunity to underreport a share of their gross income a. The government audits with a probability $(1-p)$. If audited, the individual pays a fraction $q \in [0, 1]$ of the income not reported.

Importantly, the decisions on labor supply and underreporting are made *jointly* in this framework, making taxable income an endogenous variable. Individuals substitute across two margins: the decision to take risk or no risk as well as the decision between leisure and labor supply.[9]

Individual preferences over pre-tax income $l \geq 0$ and post-tax income $c \geq 0$ can be represented by an additively separable, twice differentiable expected utility function of the form $E[u(c) - v(l)]$, where u', v', $v'' > 0 > u''$. The post-tax income is therefore given by

$$c = \begin{cases} (1-t)l + tal, & \text{with probability } p \\ (1-t)l - qal, & \text{with probability } 1-p \end{cases}$$

.

Given (t, q, p), the utility maximization problem is therefore to choose (l, c, a) to maximize $E[u(c) - v(l)]$ subject to the constraints on c, or, substituting for c, to maximize $U(l, a) = p \cdot u((1-t)l + tal) + (1-p) \cdot u((1-t)l - qal) - v(l)$.

The labor supply $l(t, q, p)$ and share of underreported income $a(t, q, p)$ are implicitly characterized by the following FOCs (u'_- denotes marginal utility when audited, u'_+ when not audited):

$$U_l := p(1 - t + ta)u'_+(\cdot) + (1-p)(1 - t - qa)u'_-(\cdot) = 0 \tag{1}$$

$$U_a := ptu'_+(\cdot) - (1-p)qu'_-(\cdot) = 0 \tag{2}$$

The revenue generated by the government is given by:

$$R(t, q, p) = (t - pta + (1-p)qa)l(t, q, p) \tag{3}$$

Therefore:

[9] Cowell (1985) discusses a very similar framework with formal and informal labor markets. He shows how substituting along these two margins only yields ambiguous predictions since "all sorts of behavior could be consistent with rational expected utility maximization."

$$\frac{\partial R}{\partial t} = (t - pta + (1-p)qa)\frac{\partial l}{\partial t} + l(t,q,p)\left[-pt + (1-p)q\right]\frac{\partial a}{\partial t} \tag{4}$$

While it is challenging to make analytical statements, it is principally possible to use the implicit function theorem to examine the effects of changes in t, q, and p on a and l (see Weiss, 1976). With the change in expected wealth and the introduction of uncertainty into final wealth due to tax evasion, there are conflicting effects at play:

Since the expected income for a given level of effort is raised, the marginal utility of wealth decreases, which in turn decreases labor supply. However, the introduction of uncertainty into final wealth due to tax evasion can have opposing effects on labor supply, depending on the shape of absolute risk aversion. If absolute risk aversion decreases with wealth (meaning that people are less risk-averse as they become wealthier), labor supply will increase with the opportunities to evade taxes. Conversely, if absolute risk aversion increases with wealth (meaning people become more risk-averse as they get wealthier), labor supply will decrease. Intuitively speaking, as individuals become wealthier, they are less afraid of taking risks. As the potential benefits of tax evasion (retaining more of their income) might motivate such individuals, this could lead to an increase in labor supply. Moreover, if these high-productive individuals anticipate a penalty for tax evasion, they may further increase their labor supply as a hedge against this potential loss of income. Empirical evidence is needed to determine which effect dominates in reality.

This theoretical framework provides us with three directly testable predictions:

Hypothesis 1. *With the incentive to cheat, tax evasion increases, i.e.,* $\frac{\partial a}{\partial t} > 0$, $\frac{\partial a}{\partial q} > 0$, *and* $\frac{\partial a}{\partial p} > 0$.

Hypothesis 2. *Labor supply will increase with the opportunities to evade taxes, i.e.,* $\frac{\partial l}{\partial t} > 0$, $\frac{\partial l}{\partial q} > 0$, *and* $\frac{\partial l}{\partial p} > 0$.

Based on the mechanism proposed by Weiss (1976), the confirmation of the first two hypotheses constitute the necessary conditions to test our third hypothesis, the centerpiece of our paper:

Hypothesis 3. *The opportunity to evade increases overall tax revenues, i.e.,* $\frac{\partial R}{\partial t} > 0$, $\frac{\partial R}{\partial q} > 0$, *and* $\frac{\partial R}{\partial p} > 0$.

3 Experimental Design

The goal of our experimental design is to mirror the basic framework proposed by Weiss (1976) in order to assess its counterintuitive implication: with reduced enforcement, overall tax revenues might increase. For this, we exploit the advantage of a controlled, abstract online-lab experiment, focusing on the mechanisms suggested in the theoretical debate.

Specifically, we operationalize the determinants *labor supply* and *tax evasion* in a deliberately abstract setting: Participants were able to earn income from a real-effort task on a piece-rate basis. For this income taxes were due (i.e. as a "fee"). Whereas the control group had no choice but to pay the fee, the treatment group was able to invest (parts of) their income into a lottery to avoid paying those fees. Since our participants were able to freely decide upon the number of tasks they are willing to perform, we were able to determine the differential labor supply and resulting tax revenue between the two treatment groups. It is important to note that the instructions strongly emphasized that payment would only be made if all the indicated tables were counted correctly, and the code that qualified the payment was displayed only then. As a result, participants had a strong incentive to declare their true willingness to work. We also elicited risk preferences using a version of the Eckel and Grossman (2002) method. The following paragraphs provide a more detailed description.[10]

3.1 Experimental Environment: A Real (Online) Labor Market

The empirical literature on tax evasion faces the fundamental issue of missing data on different levels: Firstly, the evasion opportunities themselves are hardly quantifiable since these are often situated in highly complex legislative settings. Accordingly, changes of these opportunities are even more obscure. Further, data on criminal behavior is, by nature, very elusive. Moreover, even the legal part of determinants is not straightforward to measure: Overall tax revenues are subject to constant legal changes and, on top of that, prone to macroeconomic cycles. Similarly, information on tax enforcement itself is not publicly disclosed. Even though studies roughly estimate actual audit probabilities, these are highly dependent on the specific sub-group of income earners as well as the sophisticated enforcement strategies of governmental agencies. Besides that, individual perceptions of enforcement might highly diverge from real probabilities and, therefore, are even harder to measure. Finally, reversed causality poses a problem since the level of enforcement likely is not exogenous: it responds to the level of overall tax revenues and vice versa. As a consequence, labor supply elasticities and the resulting revenue implications are difficult to estimate and might not be interpreted in a causal manner. Thus, it is close to impossible to answer the question at hand based on observational data (Slemrod and Weber, 2012). Therefore, our research question calls for a tightly controlled experiment.[11]

Such a truly randomized experiment generates reliable data on individual decisions on labor supply as a response to the opportunity to evade. The typical high "internal validity" allows us to isolate the specific mechanisms and derive causal statements on the revenue implications of evasion incentives. Against the background of the rather theoret-

[10] For exact wording and screenshots, please refer to Appendix D.

[11] Indeed, this method is widely employed in the literature on tax evasion (Alm and Malézieux, 2020).

ical motivation of our research question, we used a "neutral" frame and fully "gamified" our experiment (a common technique as an alternative to a "loaded" frame, see Alm and Malézieux, 2020). This way, we deliberately abstract from many real-life aspects, canceling out confounding factors like framing, lying aversion, moral costs or social preferences (as discussed in e.g. Slemrod, 2007; Kirchler et al., 2008).

In particular, we exploited the advantages of a real (online) labor market for our experiment; Amazon's Mechanical Turk (MTurk). MTurk is an online platform, on which usually companies post relatively simple and quick tasks (these tasks are called "Human Intelligence Tasks", HITs). These tasks are mostly repetitive like transcribing data, classifying images, transcribing audio clips, etc. (Horton et al., 2011; Berinsky et al., 2012; Paolacci et al., 2010; Mason and Suri, 2012). Recently, social scientists established this platform as a frequent subject pool for conducting experiments.[12] Multiple studies have shown that the data obtained on MTurk is as reliable as data obtained via traditional methods.[13]

For the purposes of our study, MTurk presents a number of significant advantages compared to a standard lab setting. First, MTurk samples tend to be more representative of the US population than typical student samples: these samples are usually more diverse in age, ethnicity, education and geographical location (Difallah et al., 2018; Buhrmester et al., 2011; Berinsky et al., 2012; Paolacci et al., 2010). Second, peer effects can be excluded as participants have no way of meeting the other participants. Subjects' anonymity can be sufficiently ensured as only their anonymized MTurk-ID is collected. Third, and most strikingly, experiments embedded in online labor markets present a particularly useful environment for real-effort tasks, as subjects in online experiments face real opportunity costs. With each participant being free to decide on the amount of time spent on the experiment herself, we are able to elicit the actual willingness to work. Appearing to an appointment in a physical lab with a predetermined time frame yields "sunk costs" that would motivate participants to work below their reservation wage otherwise. Exploiting the open labor market characteristic of Amazon Mechanical Turk, respondents' participation in the experiment is only granted as long as the payment scheme is perceived as profitable.[14] This allows us to overcome a common problem of labor supply experiments in labs and elicit credible preferences in labor supply.[15]

Nonetheless, leaving the lab and recruiting from a more general population on the internet also bears a couple of risks. In particular, non-US based MTurkers using Virtual Private Servers (VPSs) or automated scripts ("bots") have appeared to cause a decline

[12] For example: Suri and Watts (2011); Peysakhovich et al. (2014); Rand et al. (2014); Mao et al. (2017); Jordan et al. (2017).

[13] See among others: Arechar et al. (2018); Horton et al. (2011); Berinsky et al. (2012).

[14] Our average hourly payment was calibrated very carefully to the common average payment level on MTurk to prevent anomalous labor supply responses, see Section A.2.

[15] Related studies that implemented real effort tasks in true lab-settings only documented small labor supply responses (e.g. Doerrenberg and Duncan, 2014).

in data quality.[16] We implemented a couple of measures and checks to reduce this risk as far as possible: As is common in practice, only US-based workers, verified through IP addresses in MTurk, with an average approval rate of 95%, and at least already 500 successfully completed tasks were allowed to take part in our experiment.[17] We further implemented basic measures such as limiting the visibility of our survey to participants who signed up at MTurk with a US address and asking to confirm participants' US residency in the consent form. As a "gate-keeper" and to double check the self-indicated location, we used a third-party web servicewhich identified participants using a tool to mask their location outside the US (i.e. VPS, VPN or proxy). These participants were automatically excluded before they could enter our experiment. Next, participants had to pass a captcha-test that identifies non-human users on the first page. Subsequently, we designed an attention check which visually resembled a typical choice set between six different games, similar to the Eckel Grossmann task (Eckel and Grossman, 2002). Here, we wrote in a short text to just select the third game if they attentively read the instructions. All subjects failing this task were either not reading the instructions carefully or potentially bots. Failing this task led to the direct exclusion of the experiment. Further, to consider the possibility of participants using automated means to process our real effort task (e.g. "optical character recognition" software) we examined the average time needed for counting a single table per participant. Assuming that such a software would solve such a task in a fraction of a second, we do not find any indication for such tools.[18] Finally, we prevented workers from participating in our study more than once.[19]

3.2 Labor Supply

The task: To operationalize one of our key determinants, labor supply, we utilized the "counting-zeros" real effort task, originally applied by Abeler et al. (2011). This task asks participants to count the numbers of zeros in a 10 x 15 digit table randomly filled with 150 zeros and ones. We included this task not only because it is very monotonous and tedious and therefore includes positive costs of effort, whereas intrinsic motivation is (largely) excluded. It is also artificial with clearly no value to the experimenter so that subjects would not anticipate higher payments or ratings by the experimenter with

[16] Recent studies (e.g. Kennedy et al., 2018; Stokel-Walker, 2018) have shown that the subject pool on MTurk potentially has issues with bots, non-US based workers with poor English skills or simply inattentive participants.

[17] Requesters can review the work done by MTurkers and decide to approve or reject the work. Approved work is paid as indicated in the contract and rejected work is not paid. Hence, higher approval rates of workers indicate a higher quality of work.

[18] One average participants needed more than 70 seconds per table. Only two participants needed less than 10. For more information see Figure 8.

[19] Respondents had to enter their unique worker ID on the first page before they were able to start the survey. Further participants received a password to submit to MTurk only at completion. We clearly stated that any violation would be penalized by rejecting the HIT which would result in a significant reputational loss for workers on MTurk.

different levels of effort. Finally, it is mostly independent of ability with no mathematical or motoric skills necessary (Abeler et al., 2011, p.473). For this task, our subjects were paid on a per-piece basis and they were only able to proceed if the right number of zeros was entered.

The decision: Based on these rules, the participants were asked to indicate the number of tables they were willing to work on. In detail, each participant had to make eight different work decisions: for each of the eight different payment schemes, they had to indicate their personal willingness to work. In four "low wage" cases, participants earn $0.12 per table, and in another four "high wage" cases $0.25 per table. Within each of these two wage levels, participants have to pay either 20%, 40%, 60% or 80% of their gross income in fees to the experimenter. Both the order of these two payment blocks as well as the order of tax rates were randomized. After stating the number of tables they would be willing to work on in each of these eight scenarios, respondents were required to actually perform only one randomly picked payment scheme – a fact which was made clear to the respondents before taking their decisions. This way, we were able to elicit honest labor supply responses for all potential tax rates in a truly *incentive compatible* way: Each decision was relevant. If a participant indicated her willingness to work on e.g., 60 tasks for a scenario, and this scenario is then randomly picked, then they would actually needed to work on the indicated number of tables, which in this example would take approximately one hour. Because the payment was processed only when the participant reached the end of the experiment (i.e., after solving all the indicated tables), a non-fulfilled labor supply indication would have inevitably resulted in a total payment of $0.[20]

Table 1: Overview of the Eight Different Payment Schemes.

Payment Block	Tax Rates in %			
Low: $0.12	20	40	60	80
High: $0.25	20	40	60	80

Note: Each participant was faced with four scenarios of low payment and four scenarios of high payment per table. Both the order of these two payment blocks as well as the order of tax rates were randomized.

3.3 Treatment Variation: Tax Evasion

This aforementioned payment scheme was subject to our between-subject treatment variation. Specifically, our design comprises two main treatments in which the opportunity

[20]Further, this payment scheme was structured against the background of the typical payment on MTurk. Specifically, incentives were structured such that there is sufficient room for labor supply responses: payments were better than typically in some situations; worse in others. For further details see Section A.2.

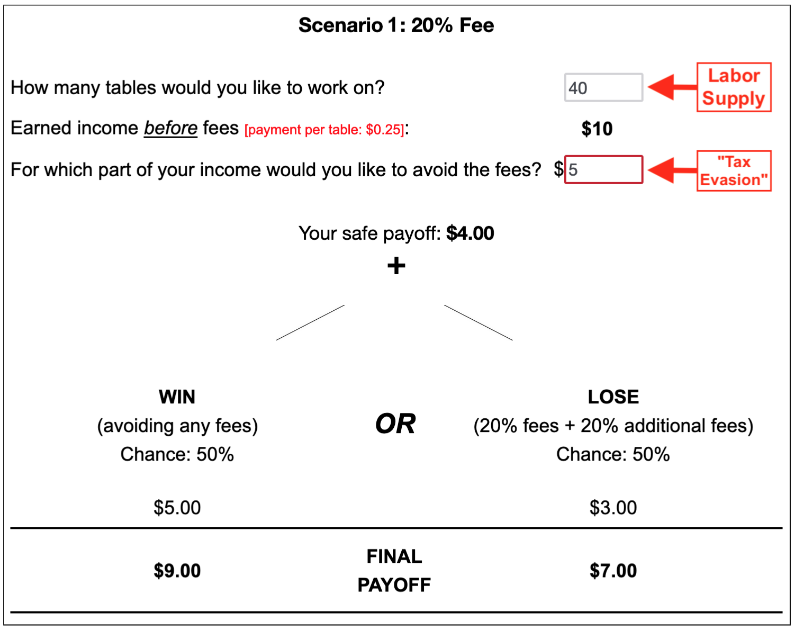
to evade is either given or not.

The Evasion-treatment: Participants jointly decided upon their labor supply and the amount evaded from their resulting before-tax income in the Evasion-treatment. In particular, participants were free to invest (parts of) their gross income into a lottery. The lottery was represented by a fair coin toss with a 50% chance to win, and a 50% chance to lose. If participants won, they avoided paying any fee on the invested amount; if they lost, they had to pay the fee on the invested amount plus an additional fee of 20% on the income invested.

These parameters were deliberately chosen: The 50% audit probability maximizes participants' comprehension. As the experiment aims to assess the theoretically suggested mechanism, we tried to prevent other confounding perceptual biases induced by e.g. small probabilities. The 20% penalty, on the other hand, provided us with a clear baseline scenario, in which risk-neutral participants were indifferent between investing into the lottery and not investing at all. Moreover, 20% constitutes a fairly realistic value.[21] To reduce the cognitive load of this decision to a minimum, we designed an interactive decision tree for this screen: The first input field asked for the number of tables the participant is willing to perform. Right below, the corresponding income before fees is calculated in real-time. The second input field then asked which part of this income the participant would have liked to avoid the fees. Based on this, the participant received immediate feedback on both potential payoffs in case of winning (i.e. paying no fees on the amount invested) or losing (i.e. paying the fees plus an additional fine on the amount invested) the lottery. Thereby, the participant was invited to play around with the two inputs while the potential payoffs are recalculated in real-time on-screen.

[21] E.g. in the US: the penalty amounts to 20% - 40% on the amount understated. See: https://www.law.cornell.edu/uscode/text/26/6662 (13.07.23)

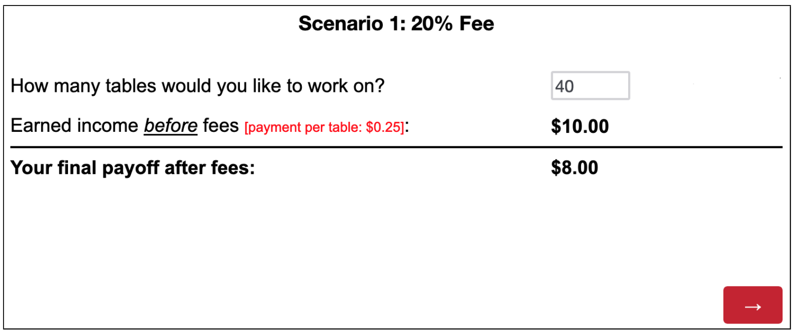
Figure 1: Screenshot of the Evasion-Treatment.



Note: Important: The red arrows and boxes were *not* part of the screen. Here, they only point out how we operationalized "labor supply" and "tax evasion" in our experiment.

The NoEvasion-treatments: In the NoEvasion-treatments,[22] subjects were only able to decide upon their labor supply, without any opportunity to avoid the fees. Based on the applicable fee, the participant received immediate on-screen feedback about the final payoff after fees. Again, participants were invited to play around with their input (i.e. number of tables) with the final payoff calculated in real-time on-screen.

Figure 2: Screenshot of the NoEvasion-Treatment.



[22] We additionally include a **NoEvasion-Lottery**-treatment. Here, subjects decided upon their labor supply and the income evaded, which directly mirrored the decision tree in the NoEvasion-Lottery-treatment. However, in this treatment tax evasion led to punishment with 100% certainty, which is unambiguously explained both in the instructions and on-screen (see screenshot in Appendix 25). This treatment was only designed to ensure the treatment effect not being driven by the longer instructions in the Evasion-treatment, the availability of a second choice or the reduced cognitive load in the NoEvasion-NoLottery-treatment. Our expectation of no systematic, significant differences to the NoEvasion-NoLottery-treatment was confirmed by the data (see Appendix C). Therefore, we pool both NoEvasion-treatments in the remainder of the paper.

3.4 Key Dependent Variable: Expected Tax Revenue

We are able to calculate the expected tax revenue for each individual ($\mathbb{E}(R)_{i,t}$) by employing the following equation (Weiss, 1976):

$$\mathbb{E}(R)_{i,t} = h_i \cdot w \cdot t - e_i \cdot (0.5 \cdot t - 0.5 \cdot 0.2) \tag{5}$$

where h indicates the labor supply (i.e. how many tables subject i indicated to work), w is the wage per table (i.e. \$0.12 or \$0.25), t is the tax level (i.e. 20%, 40%, 60% or 80%) and e_i is the income attempted to evade. With an audit probability of 50% and a penalty of 20%, the incentive to evade increases with increasing tax rates since the difference to the expected income increases. The predictions for risk-neutral, revenue maximizing participants in this setting are straight forward: While being indifferent in the 20% tax scenario, income maximizing individuals always invest all of their pre-tax income into the lottery. Or in other words; with every investment into the lottery, the government loses revenue in expectation. For the NoEvasion-treatments the term e_i is, by definition, zero.

4 Data

4.1 Organization

Our study was implemented between March and April of 2019.[23] During this period, the link to our study was called 1628 times by potential experimental subjects. 253 of those subjects used either a mobile device or a proxy server and were not allowed to take part in this study to ensure attentiveness and to exclude bots. Of the remaining 1375 subjects, 106 subjects tried to do the study several times (most of these subjects failed the attention check and tried to redo the experiment nevertheless). Further 61 subjects failed the attention check and were not allowed to continue. 163 subjects stopped the study before coming to the labor task and were hence dropped from the analysis as these subjects did not make all relevant choices.[24] Thus, overall we have 996 subjects across our treatments, with 510 subjects in the Evasion-treatment and 486 in the NoEvasion-

[23] For this experiment we obtained approval (#3tfXJpHE) by the German Association for Experimental Economic Research e.V. (GFEW) in advance.

[24] We further excluded all subjects who systematically invest into the lottery when it makes no sense (49 subjects), i.e., all those subjects who invested into the lottery for all tax levels and both wages when the detection was 100% (i.e., in the NoEvasion-Lottery-treatment). The best explanation for this behavior is confusion which is backed by 1) the fact that subjects who invest into the lottery in the NoEvasion-Lottery-treatment needed significantly ($t(64.1)$= -4.3, $p \leq 0.001$) and substantially more approaches to answer the control questions (subjects not investing into the lottery needed $M = 0.83$ ($SD = 1.05$) approaches while subjects investing into the lottery needed $M = 1.67$ ($SD = 1.28$) approaches) and 2) subjects indicated so in the open answer comment-space where participants explicitly said that they were confused by the option of investing in a sure loss-lottery. Excluding all subjects who invested instead resulted virtually in the same results. Since, as we expected, the two NoEvasion-treatments do not show any systematic differences from each other, we pool the two NoEvasion-treatments hereafter.

treatments.[25]

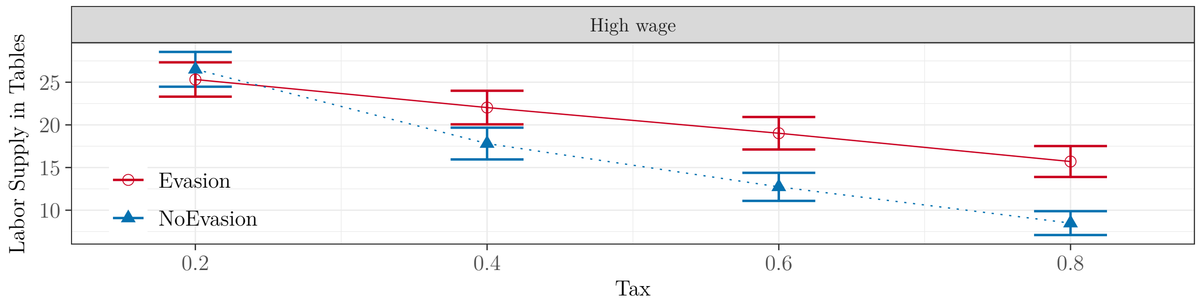
On average, participants needed 38 minutes to finish our experiment, and earned a respective hourly wage of $8.88.[26]

5 Results

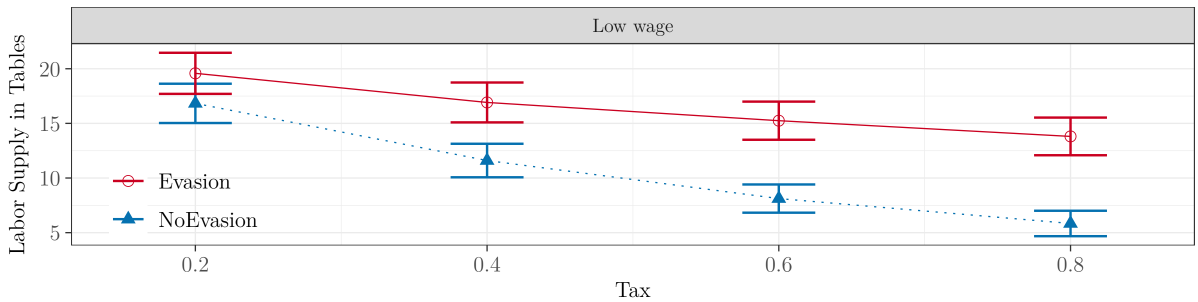
The main goal of this study is to investigate whether the option to evade taxes can increase expected tax revenue. Before coming to the main part, we will first have a look at the labor supply across treatments and a brief look at the tax evasion decisions.

5.1 Labor Supply

Figure 3: Labor Supply as a Function of Tax.



(a) Labor supply in the high-wage situation as a function of tax with 95% confidence intervals.



(b) Labor supply in the low-wage situation as a function of tax with 95% confidence intervals.

Note: Blue, dashed lines represent the NoEvasion-treatments, while red, solid lines represent the Evasion-treatment. The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals.

Concerning the labor supply, we see very clearly that incentives work. Under the most profitable condition (high wage and low tax) only 7% of all participants decided not to work. Under this condition participants on average were willing to work on 25.90

[25]NoEvasion-Lottery-treatment: 206 participants, NoEvasion-NoLottery-treatment: 280 participants.
[26]For further details see Section A.2.

tables.[27] However, under the worst condition (low wage and high tax) already 34% of all participants decided not to work. Under this condition participants on average were willing to work only on 9.92 tables. Participants also supplied significantly more work under high wages (M = 18.50; SD = 19.02) compared to low wages (M = 13.56; SD = 17.43), (t(995)= 16.1, $p \leq$0.001).

The average labor supply – aggregated over all tax levels and the two wage levels – was the highest in the Evasion-treatment with 18.45 tables, compared to the labor supply in the NoEvasion-treatments, were participants worked on 13.49 tables. On average participants in the Evasion-treatment worked 37% more compared to the NoEvasion-treatments. This difference between the Evasion-treatment and the NoEvasion-treatments is substantial and highly significant, (t(958.8)= -4.5, $p \leq$0.001).

Figure 3a depicts the labor supply in the two treatments for high wages and Figure 3b depicts the labor supply for low wages. Both graphs are rather similar and show a clear sensitivity towards the tax levels. Under the NoEvasion-treatments the labor supply is decreasing almost linearly with the tax level. Under the Evasion-treatment we can observe a similar trend but, more importantly, we can also very clearly see that the labor supply is less sensitive to the tax levels.

To investigate this relationship further we use the following mixed effects model of labor supply to estimate the treatment effects:

$$\begin{aligned} LS_{i,t,w} =& \beta_0 + \beta_1 \cdot \mathbb{1}_{\text{Evasion}} + \beta_2 \cdot t + \beta_3 \cdot \mathbb{1}_{\text{Evasion}} \cdot t + \epsilon_i + \epsilon_{i,t,w} + C_M \qquad (6)\\ C_1 =& 0 \\ C_2 =& C_1 + \beta_4 \cdot Risk_i + \beta_5 \cdot t \cdot Risk_i + \beta_6 \cdot Risk_i \cdot \mathbb{1}_{\text{Evasion}} + \beta_7 \cdot t \cdot \mathbb{1}_{\text{Evasion}} \cdot Risk_i \\ C_3 =& C_2 + \beta_X \cdot X \end{aligned}$$

where $LS_{i,t}$ represents the labor supply of subjects i for tax t under wage w with $i \in \{1, \ldots, n\}$, $t \in \{.20, .40, .60, .80\}$, and $w \in \{\text{High}, \text{Low}\}$. $\mathbb{1}_{\text{Evasion}}$ denotes a dummy with value one if the participants are in the Evasion-treatment, i.e. participants can evade their income, and zero if the participants are in the NoEvasion-treatments. t denotes the tax-level-effect, with $t \in \{.20, .40, .60, .80\}$. To account for the nested structure of the data we included ϵ_i as the random effects of the individual i. $\epsilon_{i,t,w}$ is the residuals. $Risk_i$ indicates the elicited risk preferences of subject i with higher values indicating more risk-loving. X is a vector of further control variables including age, gender, ethnicity, income, party affiliation, employment status, education, and hours spent on online work. C_1, C_2, C_3 indicate regression models with different sets of control variables.

Table 2 reports the estimates of the split regression by wage. Under a 20% tax,

[27] An apparent discrepancy between the experiment and the theory is that we measure labor supply in number of tasks whereas the theory employs time of work. Our results show clearly that these measures are empirically equivalent (see Appendix A.3 for a detailed analysis).

the Evasion and the NoEvasion-treatments do not differ, in line with our predictions. As subjects in the Evasion-treatment should be indifferent between evading and not-evading and hence, no treatment effect is expected. More interestingly, we can see that the labor supply is significantly and substantially less sensitive to tax-increases under the Evasion-treatment compared to the NoEvasion-treatments. Further, we see that risk has no influence under low wages but does have an influence under high wages. The influence of risk preferences, however, is not substantially different between the Evasion and the NoEvasion-treatments. Importantly, all results are also robust to the inclusion of controls.

Table 2: Linear Mixed-Effects Model of Labor Supply.

	Labor Supply (in tables worked)					
	Low-Wage			High-Wage		
Constant (20% Tax & NoEvasion)	16.07*** (0.82)	15.30*** (1.76)	19.10*** (3.41)	25.26*** (0.92)	19.74*** (1.96)	28.76*** (3.74)
Evasion	3.17** (1.15)	3.19 (2.52)	2.93 (2.51)	0.04 (1.28)	2.38 (2.79)	2.23 (2.78)
Tax (in %)	−0.18*** (0.01)	−0.20*** (0.02)	−0.20*** (0.02)	−0.30*** (0.01)	−0.24*** (0.02)	−0.24*** (0.02)
Tax x Evasion	0.09*** (0.01)	0.08** (0.03)	0.08** (0.03)	0.14*** (0.01)	0.08* (0.03)	0.08* (0.03)
Risk		0.23 (0.48)	0.16 (0.48)		1.68** (0.53)	1.54** (0.53)
Tax x Risk		0.01 (0.01)	0.01 (0.01)		−0.02** (0.01)	−0.02** (0.01)
Risk x Evasion		−0.03 (0.65)	0.07 (0.65)		−0.80 (0.72)	−0.75 (0.72)
Tax x Risk x Evasion		0.002 (0.01)	0.002 (0.01)		0.02* (0.01)	0.02* (0.01)
Controls	×	×	✓	×	×	✓
Sbj specific effects	✓	✓	✓	✓	✓	✓
Observations	3,984	3,984	3,984	3,984	3,984	3,984
Log Likelihood	−15,592.48	−15,598.12	−15,590.26	−16,287.54	−16,286.90	−16,274.08
Akaike Inf. Crit.	31,196.95	31,216.24	31,218.52	32,587.09	32,593.79	32,586.17
Bayesian Inf. Crit.	31,234.69	31,279.14	31,338.03	32,624.83	32,656.69	32,705.68

Notes: $^{\cdot}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$;

Note: Controls include age, gender, ethnicity, income, party affiliation, employment status, education, hours spent on online work, and the average time needed for solving the two sample tasks. Evasion denotes a dummy with value one if the participant was in the Evasion-treatment – participants have the opportunity to evade taxes and punishment will be met with a 50% probability – and zero otherwise. The omitted category are the NoEvasion-treatments. *Tax* denotes a one percentage point increase in the tax level. *Risk* denotes the elicited risk aversion with higher values indicating more risk-loving. Errors are clustered on the subject level, i.e., subject-specific effects do account for subject-specific heterogeneity.

Our results for the labor supply responses nicely tie in with other empirical findings and the theoretical predictions of the related literature. The respondents in our baseline treatment (i.e. the NoEvasion-treatments) exhibit strong negative labor supply responses towards increasing tax rates. The uncompensated labor supply and income elasticities in the NoEvasion-treatments ranges between 0.67 and 1.3 in the high wage scenario and are only slightly smaller in the low wage scenario (between 0.56 and 1.24). The empirical literature mirrors similar large elasticities of taxable income, however, mainly for individuals at the top-percentile of the income distribution (with elasticities ranging between 0.5 to 1.5, Chetty, 2009).[28] In our setting these high elasticities seem reasonable: the

[28]Generally speaking, the literature estimates substantially larger macro elasticities than micro elasticities of labor supply. Revising the restrictive assumptions usually employed in micro estimations, Löffler et al. (2014) estimate higher micro labor supply elasticities of about 0.6, closing the gap to macro estimations.

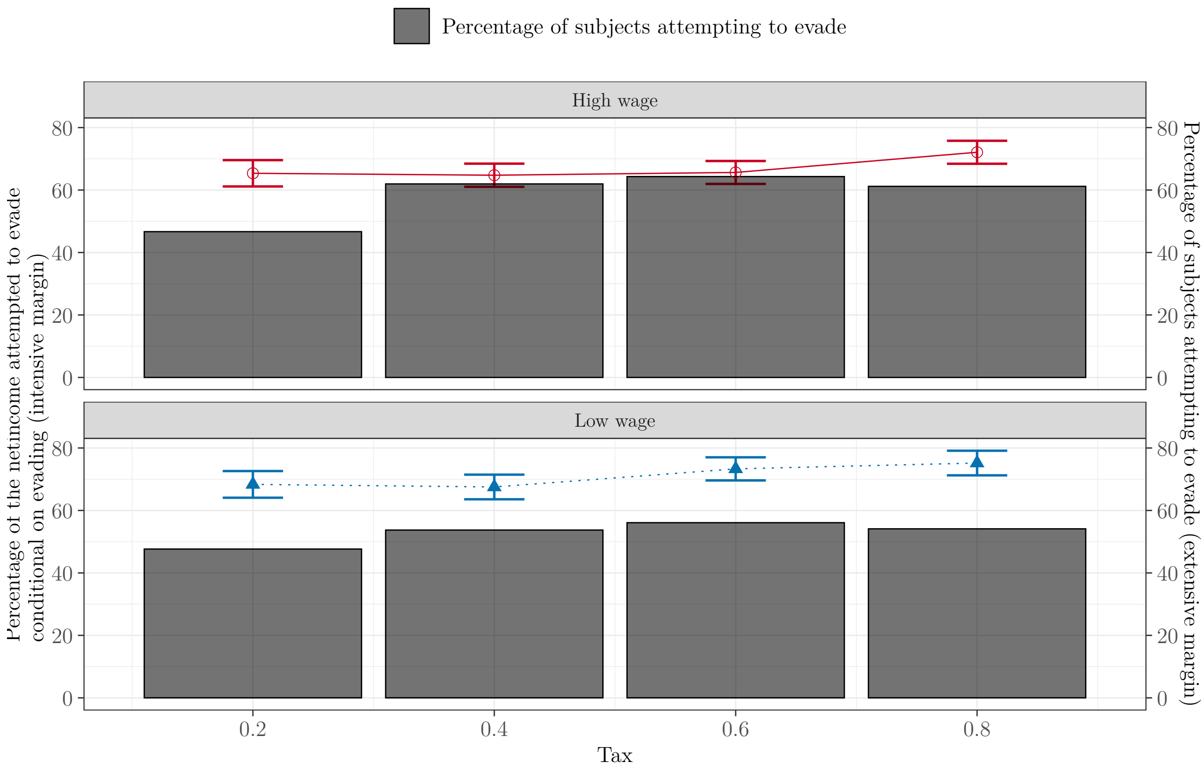
opportunity costs of the online labor market become very competitive for our high tax rates. Given the average time needed per table, the payment for the real effort task translates into an hourly payment of merely \$1.21 for the very least favorable case (low payment, highest tax rate) - against a target payment of about \$6/hr on MTurk (for a more detailed discussion see A.2).[29] In the baseline (NoEvasion-treatments), the substitution effect clearly dominates the labor supply responses across all tax rates: with each increase in the tax rate the labor supply significantly decreases, since the opportunity costs of either leisure or other tasks outside our experiment become relatively cheaper. The prominence of the substitution effect is also in line with Imbens et al. (2001) who estimated very small income effects, suggesting that the uncompensated elasticity can be approximated with the compensated elasticity of labor supply.

To summarize, we have seen that participants very clearly respond to the tax-level increase: while under the best condition 93% of participants decided to supply labor only 66% of participants decided to do so in the worst condition. Further, the labor supply decreased significantly under increasing taxes. More importantly, we observed that this labor supply decrease was significantly smaller in the Evasion-treatment. On average participants in the Evasion-treatment worked 37% more compared to the NoEvasion-treatments.

[29] See Berg (2015) and Hornuf and Vrankar (2022).

5.2 Tax Evasion

Figure 4: Tax Evasion in the Evasion-Treatment.



Note: Percentage of the net income attempted to evade conditional on evading as a function of tax with 95% confidence intervals in the Evasion-treatment. Blue, dashed lines represent the percentage of the net income attempted to evade in the low wage situation, while red, solid lines represent the percentage of the net income attempted to evade for the high wage situation (intensive margin), all conditional on evading at all. The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals. The gray bars depict the percentage of subjects attempting to evade at the separate tax levels (i.e. the extensive margin).

Concerning tax evasion, we find that most subjects (81%) decided to evade taxes in at least one setting in the Evasion-treatment. On average – aggregated over all tax-levels and the two wages situations – participants tried to evade 38.45% of their income. We can also notice that the decision to evade was very similar for the two wages.[30]

Figure 4 shows the attempted evasion as a function of the tax. We can very clearly see that the more profitable it is to evade the more participants also evade. In particular, subjects tried to evade on average 31.53% under a 20% tax while under a 80% tax subjects tried to evade 42.40% of their income, ($t(509)$= -6.3, $p \leq 0.001$) .

Looking at the extensive margin, i.e. the percentage of subjects deciding to evade, we observe that under a 20% tax on average 57% of subjects tried to evade. Under an 80% tax on average 66% of subjects tried to evade, a significant difference ($t(509)$= -3.9, p

[30] Average evasion under low wage: M = 37.66 (SD = 36.48); Average evasion under high wage: M = 39.23 (SD = 33.36), $t(509)$= 1.5, $p \geq 0.05$).

≤0.001). Looking at the intensive margin, i.e. the percentage of the net income attempted to evade conditional on evading, we notice that under a 20% tax on average 66.95% is evaded by those who evade. Under a 80% tax on average 72.54% is evaded by those who evade, which, again is highly significant ($t(509)$= -6.3, $p \leq 0.001$).

Table 3 shows a mixed effects model of the percentage of attempted evasion of participants' income as a function of the tax. Table 3 also accounts for the participants' elicited risk-preferences and further accounts for several controls. We can also observe that while higher risk attitudes seem to influence the evasion decision slightly, they do not interact with the tax level. Further, all results are robust to the inclusion of further controls.

Table 3: Linear Mixed-Effects Model of Tax Evasion.

	Tax Evasion					
		Low-Wage			High-Wage	
Constant	33.28*** (1.79)	27.56*** (4.00)	31.16** (9.94)	32.80*** (1.70)	24.91*** (3.80)	38.20*** (9.07)
Tax	0.15*** (0.03)	0.14* (0.06)	0.14* (0.06)	0.21*** (0.03)	0.23*** (0.06)	0.23*** (0.06)
Risk		1.59 (0.99)	1.65 (1.00)		2.19* (0.94)	2.31* (0.95)
Tax x Risk		0.003 (0.01)	0.003 (0.01)		−0.004 (0.02)	−0.004 (0.02)
Controls	×	×	✓	×	×	✓
Sbj specific effects	✓	✓	✓	✓	✓	✓
Observations	2,040	2,040	2,040	2,040	2,040	2,040
Log Likelihood	−10,061.83	−10,062.60	−10,049.96	−10,165.53	−10,164.84	−10,149.79
Akaike Inf. Crit.	20,131.67	20,137.20	20,129.92	20,339.06	20,341.67	20,329.58
Bayesian Inf. Crit.	20,154.15	20,170.93	20,214.23	20,361.54	20,375.39	20,413.89
Notes:	·p < 0.1;*p < 0.05;**p < 0.01;***p < 0.001;					

Note: Controls include age, gender, ethnicity, income, party affiliation, employment status, education, hours spent on online work, and the average time needed for solving the two sample tasks. The omitted category is the Evasion-treatment – participants have the opportunity to evade taxes and punishment is implemented with a 50% probability. *Tax* denotes a one percentage point increase in the tax level. *Risk* denotes the elicited risk aversion with higher values indicating more risk-loving. Errors are clustered on the subject level, i.e., subject-specific effects do account for subject-specific heterogeneity.

Our reported level of evasion in the Evasion-treatment is well in line with estimates in the evasion literature, which finds non-compliance rates in income taxation ranging from 30% to 78% (Fortin et al., 2007; Alm et al., 2009). In detail, we report an average non-compliance rate of 38% on the intensive margin, aggregated over all 8 decisions (both tax and wage levels) and on the extensive margin between 57% and 66% of subjects attempting to evade. Figure 19 in the Appendix further examines on the individual decisions to evade: the vast majority of subjects who decided to evade did so throughout all 8 scenarios (around 30% in the extensive margin) and to a large extend (around 70% of their income in the intensive margin). Furthermore, Figure 20 shows that the nature of evasion decisions (w.r.t. extensive/intensive margin) does not differ across the low or high wage level.

In summary, we have found that most subjects (81%) decided to evade taxes in at least one setting in the Evasion-treatment. On average participants tried to evade 38.45%

of their income. Further, we identified that an increasing tax rate increases the proportion of participants trying to evade.

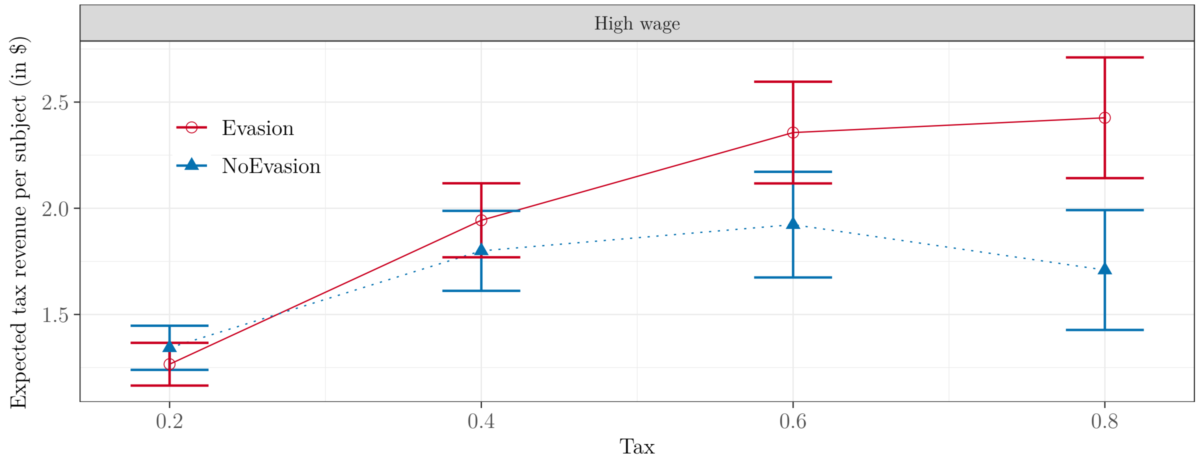
5.3 Expected Tax Revenue

We have seen that labor supply is higher if evasion is possible, however participants utilized the option to evade a substantial amount. Hence, the main question of this paper is: can the opportunity to evade still overall increase expected tax revenue? Thus, we now consider the main dependent measure of this study: the expected tax revenue. We calculate the expected tax revenue for each individual ($\mathbb{E}(R)_{i,t}$) by using the following equation:

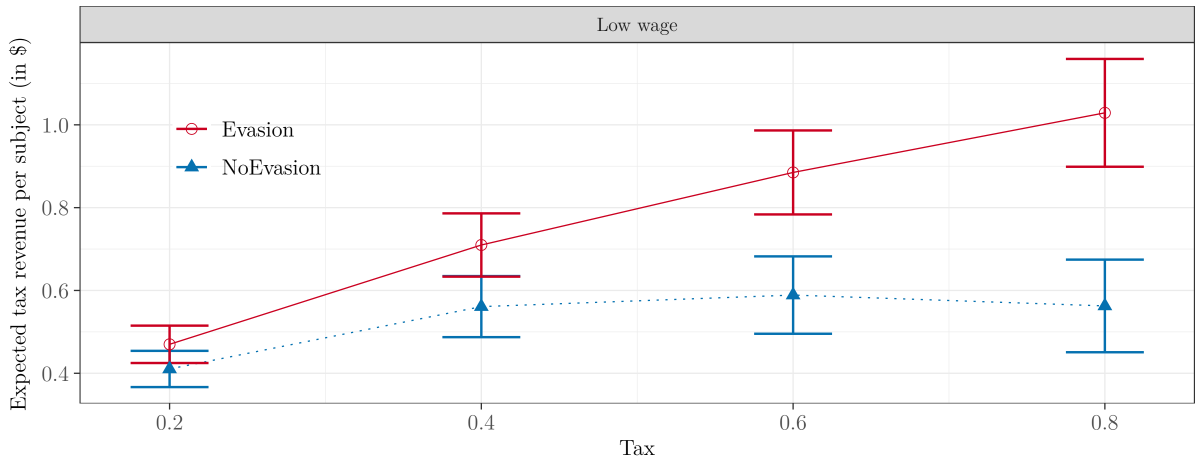
$$\mathbb{E}(R)_{i,t} = h_i \cdot w \cdot t - e_i \cdot (0.5 \cdot t - 0.5 \cdot 0.2) \tag{7}$$

where h indicates the labor supplied (i.e. how many tables subject i indicated to work), w is the wage per table, t is the tax level and e_i is the income attempted to be evaded.

Figure 5: Expected Tax Revenue.



(a) Expected tax revenue in the high-wage situation as a function of the tax with 95% confidence intervals split by treatment.



(b) Expected tax revenue in the low-wage situation as a function of the tax with 95% confidence intervals split by treatment.

Note: Blue, dashed lines represent the NoEvasion-treatments, while red, solid lines represent the Evasion-treatment. The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals.

First, we have a quick look on the two extreme scenarios: a tax rate of 20% and a tax rate of 80%. Under the nominal tax rate of 20% both treatments have an effective tax rate of 20% and do not differ in their labor supply which obviously translates into the expected tax revenue which is statistically identical between the two treatments.[31] However, the situation is very different under a tax rate of 80%. In this situation participants in the Evasion-treatment evade almost 42.40% of their income while at the same time they

[31]NoEvasion-treatments: $M = 0.88$ ($SD = 0.77$); Evasion-treatment: $M = 0.87$ ($SD = 0.79$), $t(993.5)=$ 0.2, $p \geq 0.05$. Under low wages the expected tax revenue are the following: NoEvasion-treatments: $M =$ 0.41 ($SD = 0.49$); Evasion-treatment: $M = 0.47$ ($SD = 0.52$), $t(993.9)=$ -1.9, $p \geq 0.05$). Under high wages the expected tax revenue are the following: NoEvasion-treatments: $M = 1.34$ ($SD = 1.17$); Evasion-treatment: $M = 1.27$ ($SD = 1.16$), $t(990.8)= 1.0$, $p \geq 0.05$).

increase their labor supply by 106% compared to the NoEvasion-treatments. Overall this results in a statistically highly significant and substantially higher expected tax revenue. While under the NoEvasion-treatments the expected tax revenue on average was \$1.14 it was \$1.73 in the Evasion-treatment – a highly significant difference (t(993)= -4.2, $p \leq$0.001).[32]

Let us now focus on the expected tax rate in the NoEvasion-treatments under increasing tax levels. Figure 5a depicts the expected tax revenue under the NoEvasion-treatments and the Evasion-treatment as a function of the tax rate for high wages. Figure 5b depicts the same under low wages. We can see that the expected tax revenue in the NoEvasion-treatments nicely exhibits the features of the Laffer Curve – the expected tax revenue has an inverted U-shape. In our case this is, an increase from 20% to 40% leads to a higher tax revenue. A further increase in the tax rate to 60% does not change the tax revenue significantly anymore. An even further increase in the tax rate up to 80% leads then even to a decrease in the expected tax revenue due to reduced labor supply. This picture is evident under both wage-levels.

For the Evasion-treatment the pictures look quite different: Our observations decidedly indicate that an increase in the expected tax revenue with increasing tax levels. Other than in the NoEvasion-treatments we do not find a decrease in the expected tax revenue for any increase in the tax levels (only under the high wage situation a shift from the 60% tax to the 80% tax seems to keep the expected tax revenue roughly constant).

To investigate this relationship further we use the following mixed effects model of expected tax revenue to estimate the treatment effects:

$$\begin{aligned}\mathbb{E}(R)_{i,t} =& \beta_0 + \beta_1 \cdot \mathbb{1}_{\text{Evasion}} + \beta_2 \cdot t + \beta_3 \cdot \mathbb{1}_{\text{Evasion}} \cdot t + \epsilon_i + \epsilon_{i,t,w} + C_M \\ C_1 =& 0 \\ C_2 =& C_1 + \beta_4 \cdot Risk_i + \beta_5 \cdot t \cdot Risk_i + \beta_6 \cdot Risk_i \cdot \mathbb{1}_{\text{Evasion}} + \beta_7 \cdot t \cdot \mathbb{1}_{\text{Evasion}} \cdot Risk_i \\ C_3 =& C_2 + \beta_X \cdot X\end{aligned} \tag{8}$$

$\mathbb{E}(R)_{i,t}$ represents the expected tax revenue from subjects i for tax t under wage w with $i \in \{1, \ldots, n\}$, $t \in \{.20, .40, .60, .80\}$, and $w \in \{\text{High}, \text{Low}\}$. $\mathbb{1}_{\text{Evasion}}$ denotes a dummy with value one if the participants are in the Evasion-treatment, i.e. participants can evade their income, and zero if the participants are in the NoEvasion-treatments. t denotes the tax-level-effect, with $t \in \{.20, .40, .60, .80\}$. To account for the nested structure of the data we included ϵ_i as the random effects of the individual i. $\epsilon_{i,t,w}$ is the residuals. $Risk_i$ indicates the elicited risk preferences of subject i with higher values indicating more risk-loving. X is a vector of further control variables including

[32] Under low wages the expected tax revenue are the following: NoEvasion-treatments: $M = 0.56$ (SD = 1.25); Evasion-treatment: $M = 1.03$ ($SD = 1.50$), t(978.3)= -5.3, $p \leq$0.001). Under high wages the expected tax revenue are the following: NoEvasion-treatments: $M = 1.71$ ($SD = 3.16$); Evasion-treatment: $M = 2.43$ ($SD = 3.27$), t(993.7)= -3.5, $p \leq$0.001).

age, gender, ethnicity, income, party affiliation, employment status, education, and hours spent on online work.

Table 4 reports the estimates of the split regression by wage based on dummies for the respective tax levels. Under a 20% tax the Evasion and the NoEvasion-treatments do not differ. With an increasing tax also the expected revenue increases in the NoEvasion-treatments, but only until a tax of 60%. More importantly, the expected revenue increases in Evasion-treatment for all tax levels and does significantly more so compared to the NoEvasion-treatments. We can also notice that all effects are mirrored under high wages. Further, all results are robust to the inclusion of controls.[33]

Overall – aggregated over all tax levels and all wages – the expected tax revenue is significantly higher in the Evasion-treatment ($1.30) compared to the NoEvasion-treatments ($1.01), ($t(991.4)= -3.3$, $p= 0.001$)). In fact, the expected tax revenue is on average 52% higher in the Evasion-treatment compared to the NoEvasion-treatments.[34] Thus, the answer to the main question of the paper – i.e. can the opportunity to evade overall increase the expected tax revenue – is: yes!

In summary, we observed that the expected tax revenue in the NoEvasion-treatments resembles a classic Laffer curve with peak at 60%, i.e. the expected tax revenue has an inverted U-shape relation to the increasing tax level. Further, we found, as expected, that the average expected tax revenue is indistinguishable between the Evasion-treatment and the NoEvasion-treatments under 20% tax. More importantly, we detected that the expected tax revenue in the Evasion-treatment is significantly less sensitive to the increasing tax levels. Most importantly: the expected tax revenue is, on average, higher under the Evasion-treatment compared to the NoEvasion-treatments. For a tax of 80% it is 52% higher in the Evasion-treatment compared to the NoEvasion-treatments.

[33]Obviously, the relationship between taxes and tax revenues is non-linear, which is why we use a categorical measure of taxes in this regression. In Appendix B.2 we also estimate a cubic model in Table 7. Further, we estimate a generalized additive model (GAM) to account for the non-linear relationship between taxes and tax revenues. We report the results in Table 8 We find that all the results mentioned in the main part of the paper also hold using these alternative specifications.

[34]Under low wages the expected tax revenue on average is 83% higher and under high taxes the expected tax revenue is 42% higher.

Table 4: Linear Mixed-Effects Model of Expected Tax Revenue.

	Expected Tax revenue per subject (in $)					
		Low-Wage			High-Wage	
Constant (20% Tax)	0.41*** (0.05)	0.39*** (0.10)	0.54** (0.18)	1.34*** (0.11)	1.05*** (0.24)	1.81*** (0.43)
40% Tax	0.15*** (0.04)	0.13 (0.08)	0.13 (0.08)	0.46*** (0.09)	0.33· (0.20)	0.33· (0.20)
60% Tax	0.18*** (0.04)	0.07 (0.08)	0.07 (0.08)	0.58*** (0.09)	0.48* (0.20)	0.48* (0.20)
80% Tax	0.15*** (0.04)	−0.01 (0.08)	−0.01 (0.08)	0.37*** (0.09)	0.18 (0.20)	0.18 (0.20)
Evasion	0.06 (0.06)	0.06 (0.14)	0.04 (0.14)	−0.08 (0.15)	0.06 (0.34)	0.04 (0.34)
40% Tax x Evasion	0.09· (0.05)	0.11 (0.11)	0.11 (0.11)	0.22· (0.13)	0.36 (0.28)	0.36 (0.28)
60% Tax x Evasion	0.24*** (0.05)	0.32** (0.11)	0.32** (0.11)	0.51*** (0.13)	0.61* (0.28)	0.61* (0.28)
80% Tax x Evasion	0.41*** (0.05)	0.49*** (0.11)	0.49*** (0.11)	0.79*** (0.13)	0.92*** (0.28)	0.92*** (0.28)
Risk		0.01 (0.03)	0.003 (0.03)		0.09 (0.06)	0.08 (0.06)
40% Tax x Risk		0.01 (0.02)	0.01 (0.02)		0.04 (0.05)	0.04 (0.05)
60% Tax x Risk		0.03 (0.02)	0.03 (0.02)		0.03 (0.05)	0.03 (0.05)
80% Tax x Risk		0.05* (0.02)	0.05* (0.02)		0.06 (0.05)	0.06 (0.05)
Evasion x Risk		−0.0003 (0.04)	0.005 (0.04)		−0.05 (0.09)	−0.04 (0.09)
40% Tax x Evasion x Risk		−0.01 (0.03)	−0.01 (0.03)		−0.04 (0.07)	−0.04 (0.07)
60% Tax x Evasion x Risk		−0.03 (0.03)	−0.03 (0.03)		−0.03 (0.07)	−0.03 (0.07)
80% Tax x Evasion x Risk		−0.03 (0.03)	−0.03 (0.03)		−0.04 (0.07)	−0.04 (0.07)
Controls	×	×	✓	×	×	✓
Sbj specific effects	✓	✓	✓	✓	✓	✓
Observations	3,984	3,984	3,984	3,984	3,984	3,984
Log Likelihood	−4,554.29	−4,573.44	−4,592.30	−8,135.57	−8,149.04	−8,159.52
Akaike Inf. Crit.	9,128.59	9,182.89	9,238.61	16,291.15	16,334.09	16,373.04
Bayesian Inf. Crit.	9,191.49	9,296.11	9,408.44	16,354.05	16,447.31	16,542.87

Notes: ·p < 0.1;*p < 0.05;**p < 0.01;***p < 0.001;

Note: Controls include age, gender, ethnicity, income, party affiliation, employment status, education, hours spent on online work, and the average time needed for solving the two sample tasks.Evasion denotes a dummy with value one if the participant was in the Evasion-treatment – participants have the opportunity to evade taxes and punishment will be met with a 50% probability – and zero otherwise. The omitted category are the NoEvasion-treatments. *Tax* denotes the respective tax level, i.e. 20, 40, 60 or 80%. *Risk* denotes the elicited risk aversion with higher values indicating more risk-loving. Errors are clustered on the subject level, i.e., subject-specific effects do account for subject-specific heterogeneity.

5.4 Further Mechanisms

5.4.1 (Enforced) Statutory Tax Rates vs. (Equivalent) Effective Tax Rates

Most crucial for the interpretation of our results is the comparison of tax revenues in terms of effective tax rates. One could very well argue that the opportunity to evade is nothing but a factual decrease in effective tax rates (i.e. the expected value taking into account fines and audit probabilities). According to the standard framework by Mirrlees (1971), positive labor supply responses therefore come as no surprise. Importantly, the mechanism proposed by Weiss (1976) distinctively points out a revenue increasing effect of tax evasion *beyond* that would be achieved by an equivalent decrease in statutory tax rates.

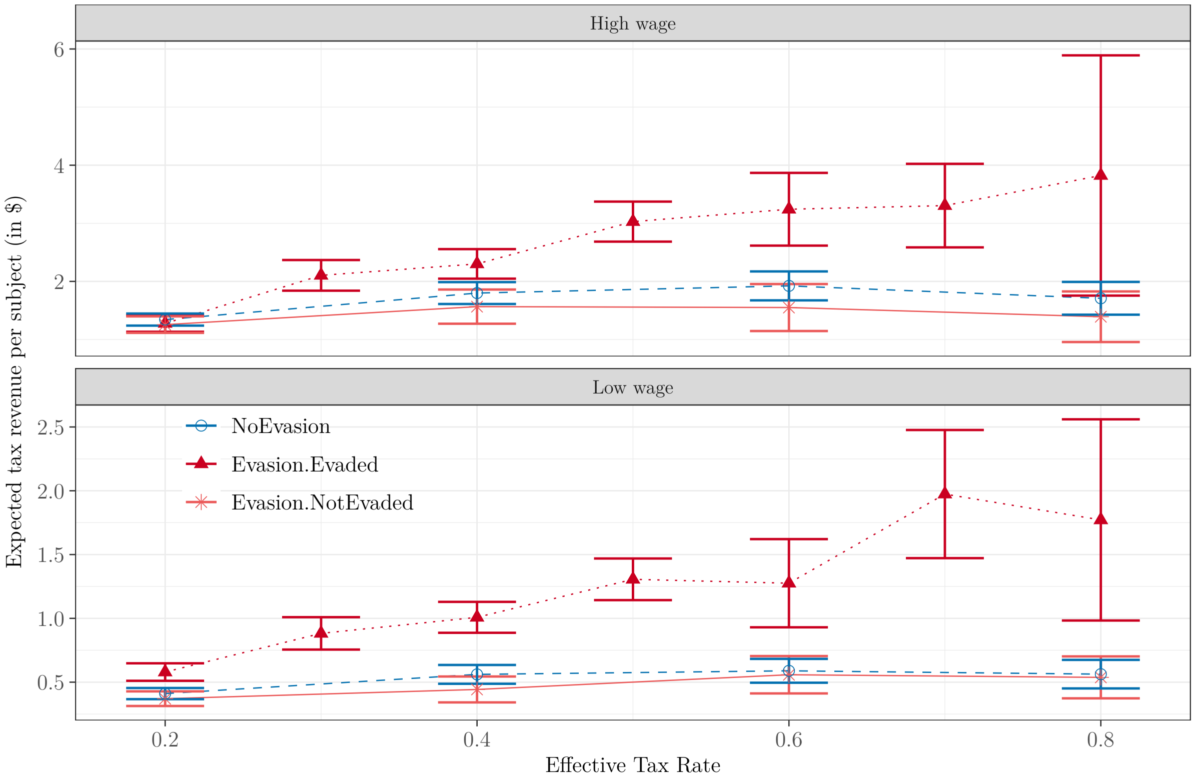
In its essence, this pronounced mechanism is already documented in the main graph 5a. With a revenue-maximizing tax rate of 60% in the NoEvasion-treatments, revenues decrease if the tax rate is decreased to 40%. However, if the treatment is changed to the

Evasion-treatment with a 60% tax-rate, tax revenues significantly increase. In case all participants would fully evade, this 60% tax-rate with the opportunity to evade translates into an equivalent 40% statutory tax rate. In this pivotal comparison between these two scenarios (40% NoEvasion vs. 60% Evasion), the effect appears even stronger. Given that the revenue-increasing effect is merely driven by the subgroup of deliberate evaders (see section 5.4.2) our estimates likely depict a lower-bound.

To examine this effect further, we first compare the statutory tax rates (i.e. 20%, 40%, 60%, 80%) with the average self-chosen effective tax rates of each respective level (resulting from the actual evasion decisions). To do so, we calculate for every subject their personal effective tax level (i.e. expected tax payment divided by the net income). For a statutory tax level of 20% the average effective tax rate is 20% – this is achieved by definition as the penalty is 20%. For a statutory tax rate of 40% the average effective tax rate is 36% while for a nominal tax rate of 60% and 80% the average effective tax rate is 50% and 64% respectively. These effective tax levels are virtually identical between the two wage levels (i.e. all effective tax levels are statistically indistinguishable between the two wage situations).

But, how do these equivalent tax rates translate into total tax revenues? In Figure 6 we illustrate the expected tax revenue in terms of the effective tax rate in comparison to the equivalent statutory tax rates of the NoEvasion-treatments. For every effective tax rate above 20%, it is clearly visible how the expected tax revenue is also higher than its statutory counterpart in the NoEvasion-treatments. Individuals who evade at a given tax rate are compiled into bins of 10%. Even further, comparing the respective higher effective rate to the lower statutory rate (e.g. 20% of the NoEvasion-treatments vs. 30% of the Evasion-treatment), tax revenues still persist to be higher. This result is especially remarkable against the background of our treatment calibration: By design, every investment into the lottery *lowered* the expected tax revenue. Nonetheless, the positive labor supply responses are so pronounced, they crowd-out the losses of tax evasion if the tax base (i.e. labor supply) would have been static - even beyond the revenue equivalent. The results very clearly show: no matter by how much the statutory tax rate is reduced, the tax revenue of a 60% tax with the opportunity to evade is never achieved.

Figure 6: Expected Tax Revenue as a Function of the Effective Tax Rate.

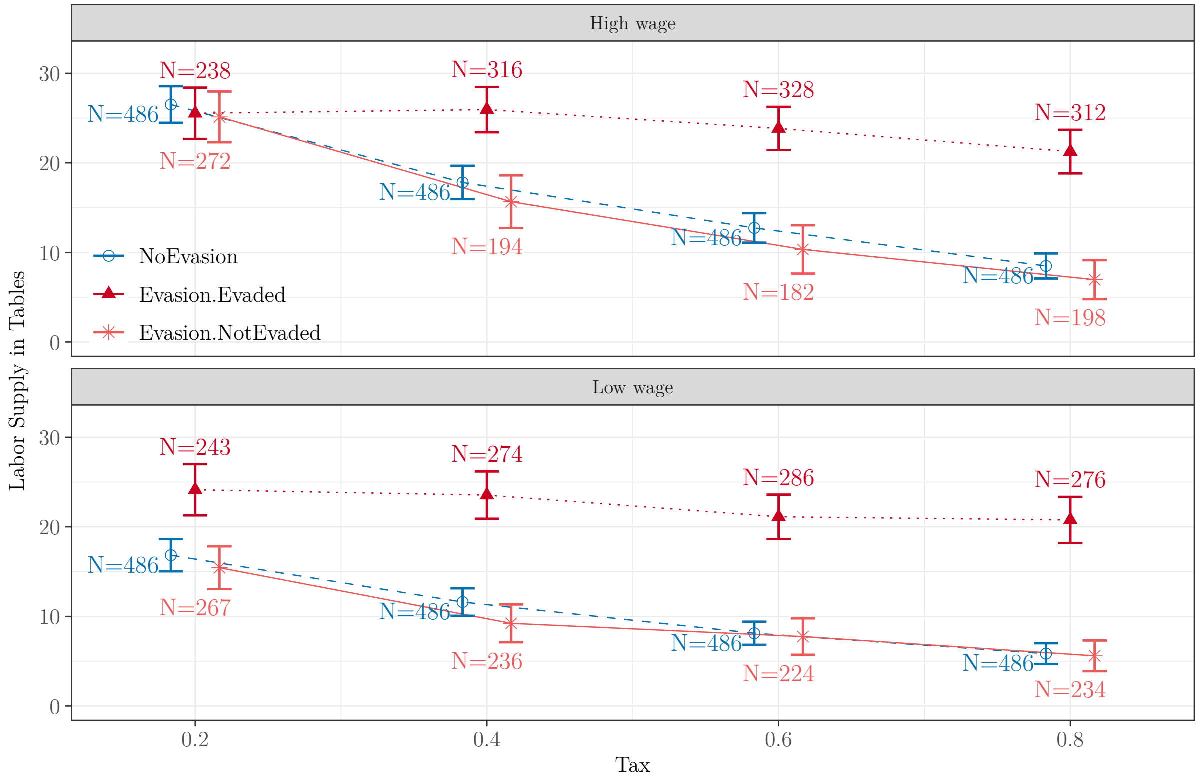


Note: Expected tax revenue as a function of the effective tax rate (exp. tax divided by the net income) with 95% confidence intervals split by treatment. Orange, dotted lines represent the Evasion-treatment where subjects did not evade; red, solid lines represent the NoEvasion-treatments; blue, dashed lines represent the Evasion-treatment where subjects evaded. The upper panel depicts the high wage situation while the panel at the bottom depicts the low wage situation. The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals.

5.4.2 Self-Selection into Evasion in the Evasion-Treatment (ATE vs. ITT)

In order to understand the composition of the labor supply responses of our Evasion-treatment group it is important to account for the possibility of self-selecting out of evasion. Our participants in the Evasion-treatment were able to discreetly pick any level of evasion between up to 100% of gross income – but also 0%. Therefore, our Evasion-treatment group also comprises individuals who deliberately did not evade. In turn, the average-treatment effect (ATE) constituted by the individuals who actually evaded in the Evasion-treatments potentially differs from the presented aggregated intention-to-treat (ITT) effect. To examine this difference, we compare the labor supply of those subjects in the Evasion-treatment who did not evade at a given tax-rate and those who evaded against the subjects who were not able to evade (NoEvasion-treatments). Figure 7 depicts for both wage-levels the labor supply as a function of the tax for these three situations. For each tax rate, about 40% of participants did not evade despite having the opportunity in the Evasion-treatment. This picture is largely driven by a subgroup of participants who actually *never* evaded throughout all of the different tax rates (ca. 22% of participants in the high wage and ca. 33% of participants in the low wage; see Figure 9). We clearly see that subjects who did not evade differ strongly from those who evaded. The deliberately non-evading subjects demonstrate a substantially lower labor supply, indistinguishable to those who could not evade (across all tax levels). Reversely, the labor supply of those who did evade was substantially higher compared to those who could not or did not want to evade.

Figure 7: Labor Supply by Evasion Decision (Extensive Margin).



Note: Labor supply of subjects by treatment and extensive margin (i.e. if a subject decided to evade within each respective tax level). Orange, solid lines represent the Evasion-treatment where subjects did not evade; blue, dashed lines represent the NoEvasion-treatment; red, dotted lines represent the Evasion-treatment where subjects evaded. The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals.

This analysis reveals how our treatment effect is driven by the subgroup of evaders. This yields the following interesting insight: it seems not to be the mere existence of the evasion opportunity driving the effect - but rather the actual evasion decision itself.

5.4.3 The Role of Other Determinants

To further explore and understand the potential determinants of who self-selected into tax evasion and who drove the large labor supply responses, we ran a couple of further analyses on personality traits in Section B of the Appendix. We do not find any systematic difference for ability[35] (see Figure 14) or other socio-economic traits like household income

[35] Proxied by the average time needed to solve the two trial tables.

(see Figure 17), importance of online work (see Figure 16) or gender (see Figure 18) on the decision to evade and labor supply. Our risk measure (similar to Eckel and Grossman, 2002), on the other hand, reveals that participants who are less risk averse (above the median) invest 4% - 10% more into the lottery (see Figure 10). However, this does not translate into higher labor supply (see Figure 12) and tax revenues.

6 Discussion & Conclusion

The findings of this project may challenge a long-standing assumption that tax evasion leads to a reduced overall tax revenue. With novel empirical insights, this project provides a more nuanced view on the effects of tax evasion on the overall tax revenue (and thus, part of the social costs of tax evasion). This project depicts, to our knowledge, the first empirical investigation of the direct relationship between the opportunity to evade and overall tax revenues. Building upon a thus far exclusively theoretical debate on potential welfare increasing effects of tax evasion initiated by Weiss (1976), we shed an empirical light on the inter-relatedness of labor supply and evasion incentives - a mechanism, which the theoretical debate deemed as empirically rather unlikely.

Given the near impossibility of cleanly exploring such a question with observational data, we exploit the advantages of a highly controlled experimental approach. Specifically, we implemented an original real effort experiment in a real (online) labor market. Our treatment group had to take a joint decision on labor supply and level of tax evasion, whereas our control group was only able to decide upon the preferred labor supply, without the opportunity to evade.

Our findings not only show significant positive labor supply responses to the opportunity to evade (increased labor supply by on average 37%). Also, the expected tax revenue significantly and substantially increased up to more than 50%. Strikingly, this effect persists when comparing effective tax rates: Lowering effective tax rates through the opportunity to evade is more efficient than simply lowering statutory tax rates, which is valid throughout all statutory tax rates above 20%. A simple back-of-the-envelope calculation shows that, for example, a 40% tax rate with complete enforcement could be replaced with a 28% tax rate with the option of tax evasion, without any loss in tax revenue. The effect is driven by the share of participants who actually evaded in the Evasion-treatment.

This empirical finding, representing the central contribution of this paper, hinges on a fundamental mechanism that should be further discussed: the differential perception of tax evasion opportunities in contrast to a simple increase of the net wage rate. In particular, the evasion decision is associated with risk and therefore perceived as more costly than an explicit decrease in tax rates, which, of course, is costless for the individual. Our findings unequivocally show that these tax reductions are indeed *not* perceived equivalent. Whereas the theoretical literature deemed the empirical relevance as highly implausible (Sandmo, 1981; Hellwig, 2007), recent literature in both the expected- and the non-expected utility framework reveal further potential behavioral underpinnings for this effect to occur. Firstly, the concept of second-order risk aversion, as elaborated by Segal and Spivak (1990), could provide a compelling explanation. If individuals perceive the risk of penalties for tax evasion as relatively small, their second-order risk aversion could manifest as risk-neutral behavior. Consequently, the potential gains from tax evasion might be deemed to outweigh the potential losses, despite inherent uncertainty. Additionally, expectations-based reference points heavily influence decision-making (Kőszegi and Rabin, 2006). If individuals perceive the potential gains from tax evasion as a new "status quo" or reference point to maintain or exceed, they might be more inclined to take the associated risk. Based on that, a novel paper by Campos-Mercade et al. (2022) implies that individuals' responses to income uncertainty are not uniformly distributed: Some individuals demonstrate an increased tolerance for risk and uncertainty, prompting a preference for potentially high-yield, high-risk strategies like tax evasion. Eventually, the concept of prudence (e.g. Kimball, 1990; Ebert and Wiesen, 2014) provides a coherent explanation for the increased labor supply as a hedge against possible losses as a self-insurance mechanism in the expected utility framework. Consequently, heterogeneity in gain-loss attitudes, together with second-order risk aversion, may account for both the general trend of uncertainty aversion and the significant subset of respondents opting for tax evasion, despite inherent risk. Despite our compelling findings and proposed mechanisms discussed, a comprehensive theory that encapsulates these dynamics in the context of tax evasion remains missing. Creating and refining such a comprehensive theory would certainly make for an interesting and valuable future project.

Our quite straightforward confirmation of a surprisingly under-studied question further constitutes a potentially fundamental contribution to the notion on welfare implications of tax evasion, i.e. the excess burden or deadweight loss of taxation. A couple of sparse works already pointed towards possible welfare increasing effects of allowing some degree of tax evasion: Chetty (2009) argues that tax evasion is not simply lost resources to society. It rather translates into transfers to others like charitable giving and thus still yields welfare increasing effects. Taking such social benefits of sheltering into account, Keen and Slemrod (2017) discuss the optimal degree of tax enforcement against the background of costly enforcement. However, our argument is somewhat more involved: even in the absence of tax enforcement costs and social benefits of tax sheltering (like charitable giving or paid penalties), the interaction between the evasion opportunity and labor supply decision adds a revenue increase beyond what would be achieved by an equivalent tax rate decrease. Simply put, the excess burden of a given tax system encouraging tax evasion is smaller than that of a system discouraging it.[36]

[36] However, this might only be the case in a situation where tax rate is suboptimal and not an efficient one, as Yitzhaki (1987, p. 134) importantly points out.

Our research design, as a standard principle of lab experiments, highly abstracts from many real-life aspects. Since our research question is rooted in a theoretical debate, this abstraction depicts an advantage to less controlled settings with a high number of confounds. On the other hand, this offers room for potential critique in terms of external validity. Against this background, three main limitations need to be discussed: First, the choice of parameters concerning audit probability, penalty and tax rates. Especially our audit probability of 50% is most likely higher than in reality. However, the main objective of our study was to maximize internal validity. That is, to reduce the cognitive load in order to elicit unbiased preferences and construct a true baseline in which participants are indifferent (i.e. 20% tax with 50% audit probability). To facilitate full comprehension, we therefore aimed to keep a fairly complex joint decision on labor supply and evasion as simple as possible. Moreover, actual audit probabilities are rather unknown to the public and, also, highly dependent on the income level as well as profession.[37] Finally, even if commonly known, the perception of these audit probabilities is likely to be biased in itself (c.f. Tversky and Kahneman, 1974). Second, we deliberately refrained from employing any type of loaded framing. That includes no moral costs of lying, no social costs of lost tax revenues (i.e. no redistribution), and no administrative costs of evasion (i.e. effort to exploit loopholes). In short, we excluded all externalities of tax evasion. Admittedly, this constitutes a strong assumption, as it is well known that individuals do not respond to sanctions as mere profit-maximizers, and even less so in social settings (Engel, 2014). Nonetheless, we argue this being a necessity in order to identify the "core" mechanism at hand. Biases like inequality aversion or guilt aversion would introduce noise to our findings, which would have prevented cleanly isolating the mechanism proposed by the theory. Moreover, even though the use of frames usually increases compliance (Alm and Malézieux, 2020), this effect is not necessarily given and could depend on the interaction with other determinants (Alm et al., 1992). Furthermore, even if externalities are introduced, the direction of their effect on tax revenues would not be straightforward: Individuals might adjust on the margin of tax evasion. However, they may also adjust on the margin of increased labor supply. Thus, we believe that a clean and internally reliable design yields a better starting point to address our research question. Future research might, however, also focus on more externally valid designs. Third, lab experiments are typically criticized for the use of small stakes. For our case, stakes were actually as close to reality as they can be: since our experiment was posted on the online labor market MTurk our participants are very likely to be part of the "gig economy". Being one of the largest platforms for online work, a majority on this platform considers MTurk "as a job" with a significant share reporting it being their primary source of income (Keith et al., 2019). With an average of 38 minutes, these workers faced real opportunity costs as they could have engaged in other income-generating tasks on this platform. If they

[37] See https://www.irs.gov/about-irs/irs-audit-rates-significantly-increase-as-income-rises (13.07.23)

lost the lottery, they could have worked for close to nothing; if they won the lottery, our payment was above average.[38] Also, our real effort task of counting zeros yields very little scope for intrinsic motivation. Thus, we do not expect our participants to participate for different than monetary reasons.[39] Finally, our estimated elasticities of taxable income match those found in observational data (Chetty, 2009). Central to our documented mechanism is the possibility to continuously adjust on the (intensive) labor supply margin. Admittedly, this is highly unlikely for most employees under regular work contracts (which are, moreover, also usually subject to third-party reporting). However, this condition does hold true for self-employed individuals and sole proprietors of small businesses, who represent more than 10% of total US tax revuenues:[40] their ETI to changes in the net-of-tax rates are substantially larger than those of salary incomes (Carroll et al., 2001; Heim, 2010). Besides that, this subgroup is even more interesting for our case since they exhibit systematically lower levels of risk aversion (Ekelund et al., 2005). Overall, we acknowledge the abstract character of our research design and the potentially resulting limitations in external validity. However, it is deliberately kept very abstract and is reduced to the critical components of the research question.

Tax evasion has many negative consequences for society: increased unfair distribution of wealth, unequal treatment of people, undermining the rule of law. However, one of the main problems typically associated with tax evasion, i.e. reduced tax revenues, might not be as obvious as typically thought. In answering the last question, this research agenda speaks only indirectly to the moral question of tax evasion. It is beyond the scope of this project to answer the moral and philosophical questions arising if tax revenue could be increased by the opportunity to evade taxes.[41] Future research will need to tackle the pressing moral and social problems associated with such a mechanism. Nonetheless, we propose a couple of potential policy implications suggested by our results: Given that governments cannot (for financial and/or moral reasons) deter all tax evasion, the question is how to most efficiently deploy fiscal capacities. Our findings suggest several recommendations. First, our effect is relatively largest for low-income workers. Thus, targeting high-income individuals is more beneficial. Second, tax enforcement is less disturbing in labor markets with low labor supply elasticities. Through this lens, already

[38] See Appendix A.2 for further analyses.

[39] In addition, it has been repeatedly shown that stake-size in these games do not substantially impact behavior. For example, Forsythe et al. (1994) and Carpenter et al. (2005) provide further compelling evidence that mean allocations in dictator games with low stakes do not differ from allocations in dictator games with high stakes. Additionally, Abeler et al. (2019) find that the amount of lying in lying games is not correlated with the size of the stakes in these games. Further, Camerer and Hogarth (1999) survey the experimental economics literature and shows that behavior is impacted mainly if tasks are incentivized. Thus, it is unlikely that our results would substantially change if we were to increase the incentives considerably.

[40] For 2015, see: `https://www.bls.gov/spotlight/2016/self-employment-in-the-united-states/home.htm`

[41] See Sandmo (1981) for an interesting discussion on the desirability of a purely utility maximizing taxpayer or a "Kantian" driven tax payer as well as the implication for horizontal equity.

incomplete enforcement for self-employed should focus on respective sectors whereas strict third-party reporting should be relaxed where positive labor supply responses are to be expected. Thus, our paper suggests novel ways for tax administrations to address the persistent challenge of tax evasion and eventually harness this issue, turning it into a strategic advantage to increase enforcement efficiency.

Acknowledgements

We thank Felix Bierbrauer, Christioph Engel, Christian Traxler, Rupert Sausgruber, Floris Zoutman, Marius Vogel, Fred Schroyen, Philipp Dörrenberg, Malte Chirvi, Jerome Olsen, Till Weber, Henrik Orzen, Arthur Seibold, Mohammed Mardan, Nina Serdarevic, Fabian Winter, Thomas Tröger for helpful comments and suggestions. We appreciate comments from the participants of the Lab Meeting at the Max Planck Institute for Research on Collective Goods in Bonn, the Mannheim-ZEW-Seminar, the IMPRS-Thesis-Workshop, the 6th Shadow Economy Conference, the 2019 ESA European Meeting, the 6th Annual MaTax Conference, the 15th Winter School on Inequality and Social Welfare Theory, the 35th EEA Congress, the 2020 ESA's Job-Market Candidates Seminar, the 2020 Meeting of the German Economic Association, the CBE Seminar as well as the NoCeT and FAIR Group at NHH. We gratefully acknowledge funding from the Max-Planck-Society. The work leading to this publication was supported by the PRIME program of the German Academic Exchange Service (DAAD) with funds from the German Federal Ministry of Education and Research (BMBF). Support by the German Research Foundation (DFG) through CRC TR 224 (Project A01) is gratefully acknowledged.

References

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2):470–92.

Abeler, J. and Jäger, S. (2015). Complex tax incentives. *American Economic Journal: Economic Policy*, 7(3):1–28.

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

Allingham, M. G. and Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of public economics*, 1(3-4):323–338.

Alm, J., Jackson, B. R., and McKee, M. (2009). Getting the word out: Enforcement information dissemination and compliance behavior. *Journal of Public Economics*, 93(3-4):392–402.

Alm, J. and Malézieux, A. (2020). 40 years of tax evasion games: a meta-analysis. *Experimental Economics*, pages 1–52.

Alm, J., McClelland, G. H., and Schulze, W. D. (1992). Why do people pay taxes? *Journal of public Economics*, 48(1):21–38.

Alstadsæter, A., Johannesen, N., Herry, S. L. G., and Zucman, G. (2022). Tax evasion and tax avoidance. *Journal of Public Economics*, 206:104587.

Alstadsæter, A., Johannesen, N., and Zucman, G. (2019). Tax evasion and inequality. *American Economic Review*, 109(6):2073–2103.

Arechar, A. A., Gächter, S., and Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, 21(1):99–131.

Baretti, C., Huber, B., and Lichtblau, K. (2002). A tax on tax revenue: The incentive effects of equalizing transfers: Evidence from germany. *International Tax and Public Finance*, 9(6):631–649.

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime*, pages 13–68. Springer.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.

Berg, J. (2015). Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers. *Comp. Lab. L. & Pol'y J.*, 37:543.

Bergeron, A., Tourek, G., and Weigel, J. (2021). The state capacity ceiling on tax rates: Evidence from randomized tax abatements in the drc.

Berinsky, A. J., Huber, G. A., and Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis*, 20(3):351–368.

Best, M. C., Brockmeyer, A., Kleven, H. J., Spinnewijn, J., and Waseem, M. (2015). Production versus revenue efficiency with limited tax capacity: theory and evidence from pakistan. *Journal of political Economy*, 123(6):1311–1355.

Blackwell, C. (2010). A meta-analysis of incentive effects in tax compliance experiments. *Developing alternative frameworks for explaining tax compliance*, 59:97.

Block, M. K. and Heineke, J. M. (1973). The allocation of effort under uncertainty: the case of risk-averse behavior. *Journal of Political Economy*, 81(2, Part 1):376–385.

Bönke, T., Schröder, C., and Jochimsen, B. (2017). Fiscal equalization and tax enforcement. *German Economic Review*, 18(3):377–409.

Bott, K. M., Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2019). You've got mail: A randomized field experiment on tax evasion. *Management Science*.

Brito, D. L., Hamilton, J. H., Slutsky, S. M., and Stiglitz, J. E. (1995). Randomization in optimal income tax schedules. *Journal of Public Economics*, 56(2):189–223.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2011). Amazon's mechanical turk. *Perspectives on Psychological Science*, 6(1):3–5.

Bustos, S., Pomeranz, D., Serrato, J. C. S., Vila-Belda, J., and Zucman, G. (2022). The race between tax enforcement and tax planning: Evidence from a natural experiment in chile. Technical report, National Bureau of Economic Research.

Camerer, C. F. and Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19(1):7–42.

Campos-Mercade, P., Goette, L., Graeber, T., Kellogg, A., and Sprenger, C. (2022). Heterogeneity of gain-loss attitudes and expectations-based reference points. Technical report, Working Paper.

Carpenter, J., Verhoogen, E., and Burks, S. (2005). The effect of stakes in distribution experiments. *Economics Letters*, 86(3):393 – 398.

Carroll, R., Holtz-Eakin, D., Rider, M., and Rosen, H. S. (2001). Personal income taxes and the growth of small firms. *Tax policy and the economy*, 15:121–147.

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, 87:43–51.

Chetty, R. (2009). Is the taxable income elasticity sufficient to calculate deadweight loss? the implications of evasion and avoidance. *American Economic Journal: Economic Policy*, 1(2):31–52.

Cowell, F. A. (1985). Tax evasion with labour income. *Journal of public economics*, 26(1):19–34.

Cremer, H. and Gahvari, F. (2000). Tax evasion, fiscal competition and economic integration. *European Economic Review*, 44(9):1633–1657.

Dave, C., Eckel, C. C., Johnson, C. A., and Rojas, C. (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty*, 41(3):219–243.

Deck, C. and Schlesinger, H. (2010). Exploring higher order risk effects. *The Review of Economic Studies*, 77(4):1403–1420.

Difallah, D., Filatova, E., and Ipeirotis, P. (2018). Demographics and dynamics of mechanical turk workers. In *Proceedings of the eleventh acm international conference on web search and data mining*, pages 135–143. ACM.

Doerrenberg, P. and Duncan, D. (2014). Experimental evidence on the relationship between tax evasion opportunities and labor supply. *European Economic Review*, 68:48–70.

Ebert, S. and Wiesen, D. (2014). Joint measurement of risk aversion, prudence, and temperance. *Journal of Risk and Uncertainty*, 48(3):231–252.

Eckel, C. C. and Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and human behavior*, 23(4):281–295.

Ekelund, J., Johansson, E., Järvelin, M.-R., and Lichtermann, D. (2005). Self-employment and risk aversion—evidence from psychological test data. *Labour Economics*, 12(5):649–659.

Emran, M. S. and Stiglitz, J. E. (2005). On selective indirect tax reform in developing countries. *Journal of public Economics*, 89(4):599–623.

Engel, C. (2014). Social preferences can make imperfect sanctions work: Evidence from a public good experiment. *Journal of Economic Behavior & Organization*, 108:343–353.

Feldstein, M. (1999). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics*, 81(4):674–680.

Fellner, G., Sausgruber, R., and Traxler, C. (2013). Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association*, 11(3):634–660.

Forsythe, R., Horowitz, J. L., Savin, N., and Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6(3):347 – 369.

Fortin, B., Lacroix, G., and Villeval, M.-C. (2007). Tax evasion and social interactions. *Journal of Public Economics*, 91(11-12):2089–2112.

Gemmell, N. and Hasseldine, J. (2014). Taxpayers' behavioural responses and measures of tax compliance 'gaps': A critique and a new measure. *Fiscal Studies*, 35(3):275–296.

Gruber, J. and Saez, E. (2002). The elasticity of taxable income: evidence and implications. *Journal of public Economics*, 84(1):1–32.

Heim, B. T. (2010). The responsiveness of self-employment income to tax rate changes. *Labour Economics*, 17(6):940–950.

Hellwig, M. F. (2007). The undesirability of randomized income taxation under decreasing risk aversion. *Journal of Public Economics*, 91(3-4):791–816.

Hornuf, L. and Vrankar, D. (2022). Hourly wages in crowdworking: A meta-analysis. Technical report, CESifo.

Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3):399–425.

Imbens, G. W., Rubin, D. B., and Sacerdote, B. I. (2001). Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players. *American economic review*, 91(4):778–794.

Internal Revenue Service (IRS) (2019). Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011–2013. https://www.irs.gov/pub/irs-pdf/p1415.pdf.

Internal Revenue Service (IRS) (2020). Internal Revenue Service Progress Update, Fiscal Year 2019. Publication 5382 (12-2019).

Jordan, J. J., Sommers, R., Bloom, P., and Rand, D. G. (2017). Why do we hate hypocrites? Evidence for a theory of false signaling. *Psychological Science*, 3(1):356–368.

Keen, M. and Slemrod, J. (2017). Optimal tax administration. *Journal of Public Economics*, 152:133–142.

Keith, M. G., Harms, P., and Tay, L. (2019). Mechanical turk and the gig economy: Exploring differences between gig workers. *Journal of Managerial Psychology*.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P., and Jewell, R. (2018). The shape of and solutions to the mturk quality crisis. *Available at SSRN*.

Kimball, M. S. (1990). Precautionary saving and the marginal propensity to consume. Technical report, National Bureau of Economic Research.

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

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., and Saez, E. (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica*, 79(3):651–692.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Lederman, L. and Sichelman, T. (2013). Enforcement as substance in tax compliance. *Wash. & Lee L. Rev.*, 70:1679.

Löffler, M., Peichl, A., and Siegloch, S. (2014). Structural labor supply models and wage exogeneity. *ZEW-Centre for European Economic Research Discussion Paper*, (14-040).

Mao, A., Dworkin, L., Suri, S., and Watts, D. J. (2017). Resilient cooperators stabilize long-run cooperation in the finitely repeated prisoner's dilemma. *Nature Communications*, 8(13800).

Mascagni, G. (2018). From the lab to the field: A review of tax experiments. *Journal of Economic Surveys*, 32(2):273–301.

Mason, W. and Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior research methods*, 44(1):1–23.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208.

Neisser, C. (2021). The elasticity of taxable income: A meta-regression analysis. *The Economic Journal*, 131(640):3365–3391.

Noussair, C. N., Trautmann, S. T., and Van de Kuilen, G. (2014). Higher order risk attitudes, demographics, and financial decisions. *Review of Economic Studies*, 81(1):325–355.

Office of Management and Budget (2013). Historical Tables: Budget of the U.S. Government, FY 2014, Table 3.2. https://www.whitehouse.gov/omb/historical-tables/.

Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5):411–419.

Peysakhovich, A., Nowak, M. A., and Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nature Communications*, 5:4939.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, 105(8):2539–69.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., and Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, 5:3677.

Saez, E., Slemrod, J., and Giertz, S. H. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of economic literature*, 50(1):3–50.

Sandmo, A. (1981). Income tax evasion, labour supply, and the equity—efficiency tradeoff. *Journal of Public Economics*, 16(3):265–288.

Schroyen, F. (1997). Pareto efficient income taxation under costly monitoring. *Journal of Public Economics*, 65(3):343–366.

Segal, U. and Spivak, A. (1990). First order versus second order risk aversion. *Journal of Economic Theory*, 51(1):111–125.

Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives*, 21(1):25–48.

Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*, 57(4):904–54.

Slemrod, J. and Traxler, C. (2010). Optimal observability in a linear income tax. *Economics Letters*, 108(2):105–108.

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

Stiglitz, J. E. (1982). Utilitarianism and horizontal equity: The case for random taxation. *Journal of Public Economics*, 18(1):1–33.

Stö, S. and Traxler, C. (2005). Tax evasion and auditing in a federal economy. *International Tax and Public Finance*, 12(4):515–531.

Stokel-Walker, C. (2018). Bots on amazon's mechanical turk are ruining psychology studies. *New Scientist*.

Suri, S. and Watts, D. J. (2011). Cooperation and contagion in web-based, networked public goods experiments. *PLOS ONE*, 6(3):1–18.

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.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131.

Wahl, I., Kastlunger, B., and Kirchler, E. (2010). Trust in authorities and power to enforce tax compliance: An empirical analysis of the "slippery slope framework". *Law & Policy*, 32(4):383–406.

Weiss, L. (1976). The desirability of cheating incentives and randomness in the optimal income tax. *Journal of Political Economy*, 84(6):1343–1352.

Yitzhaki, S. (1987). On the excess burden of tax evasion. *Public Finance Quarterly*, 15(2):123–137.

A Characteristics of Participants

A.1 Demographics

The median age of our participants was 37 years (on average = 40.16), ranging from 18 to 87 years. Most of our subjects are in the age group between 30 and 44 (46%), followed by the age group of 45 to 64 years (27%) and the age group of 18 to 29 (22%). Overall, our sample is slightly younger than the average US-American with a median age of 38.2 and with 16% of the population older than 65 years (compared to 5% in our sample).[42] In terms of gender, our sample is fairly balanced: 52% of our participants were female compared to 50.8% females in the US population. The ethnic composition is less representative: 80% of subjects are White compared to 61.3% Whites in the US population. Moreover, our participants indicated to have a higher education than the average US citizen. 60% of subjects implied to have at least a Bachelor's degree as the highest qualification compared to roughly 33% in the United States as a whole. Hence, our sample is younger, slightly more female, more white and better educated than the average US citizen. Even though our sample does not fully represent the typical American, it is substantially more diverse than the generic student sample typically used for taxation-experiments.

Table 5 investigates whether subjects' characteristics are balanced across treatments. Table 5 shows and compares all characteristics of all subjects across treatments. It is straightforward that all characteristics are evenly balanced across treatments – no treatment differs in any characteristic from the other treatments. Thus, random assignment of subjects to the treatments worked nicely.

[42] For comparison estimates see the census aggregates: https://www.census.gov/quickfacts/fact/table/US/PST045216 (13.07.23) and https://www.census.gov/data/tables/2019/demo/educational-attainment/cps-detailed-tables.html (13.07.23).

Table 5: Characteristics of Subjects in the Three Treatments.

	Evasion		NoEvasion-Lottery		NoEvasion-NoLottery		Any Sign. Differences?
	$\overline{M}$	$CI_{95\%}$	$\overline{M}$	$CI_{95\%}$	$\overline{M}$	$CI_{95\%}$	
Age	39.68	[38.60,40.76]	40.34	[38.54,42.14]	40.46	[38.95,41.97]	×
Female	00.51	[00.47,00.55]	00.51	[00.45,00.57]	00.56	[00.50,00.62]	×
HighDegree	00.69	[00.65,00.73]	00.69	[00.63,00.75]	00.65	[00.59,00.71]	×
Employed	00.75	[00.71,00.79]	00.78	[00.72,00.84]	00.71	[00.65,00.77]	×
HighIncome	00.52	[00.48,00.56]	00.55	[00.49,00.61]	00.50	[00.44,00.56]	×
Democrats	00.57	[00.53,00.61]	00.51	[00.45,00.57]	00.49	[00.43,00.55]	×
White	00.80	[00.76,00.84]	00.81	[00.75,00.87]	00.79	[00.75,00.83]	×
Risk	05.26	[05.04,05.48]	04.76	[04.43,05.09]	04.90	[04.63,05.17]	×
HoursOnlineWork	17.31	[15.88,18.74]	16.14	[14.47,17.81]	17.36	[15.95,18.77]	×
N	510		206		280		

We use two-sample t-tests to compare all characteristics. In particular, we test every treatment against every other treatment. × denotes no significant differences between any treatments on the particular dimension. ✓ denotes a significant difference in at least one of the comparisons at a 5% level on the particular dimension. We use the Benjamini and Hochberg (1995) p-value adjustment. $\overline{M}$ denotes the mean of the respective characteristic. $CI_{95\%}$ denotes the 95% confidence interval. *Age* denotes a continuous variable on the age of the participants. *Female* denotes a dummy with value one if the participant is female. *HighDegree* denotes a dummy with value one if the participant has at least a College-degree. *Employed* denotes a dummy with value one if the participant indicated to be either full time or half-time employed. *HighIncome* denotes a dummy with value one if the participant lives in a household with at least a yearly income of 75k. *Democrats* denotes a dummy with value one if the participant indicated to vote for Democrats. *White* denotes a dummy with value one if the participant indicated to be Caucasian. *Risk* denotes a continuous variable on the risk preferences indicated in the -task. *HoursOnlineWork* denotes a continuous variable on the indicated hours the participant works online per week.

A.2 Payment

The average time participants needed to finsih our experiment was 38 minutes. For this, they earned a respective hourly wage of \$8.88. Taking into account the ability of subjects (i.e. how fast subjects were able to solve the two sample tasks) we can calculate the maximum amount of tables subjects would have been able to work on in an hour. For a given hour the average number of tables subjects could work on are 50.5 [25% quantile=32.7; 75% quantile= 64.52]. Thus, the maximum payoff per hour on average in case of low wage was \$6.06 if no tax would have to be paid; the maximum payoff per hour under high wage was: \$12.63. Under low wage and 20%, vs. 80% tax the maximum payoff per hour was accordingly \$4.85 vs. \$1.21. Under high wage and 20%, vs. 80% tax the maximum payoff per hour was accordingly \$10.1 vs. \$2.53.

Looking more specifically at low ability workers (subjects needing on average more time to finish the two sample tasks than the median subject) we see that they would obtain on average a maximum payoff per hour of \$3.85 under the low wage situation and \$8.02 under the high wage situation if no taxes would need to be paid. Looking at high ability workers (subjects requiring on average less time to finish the two sample tasks than the median subject) we see that they would obtain on average a maximum payoff per hour of \$8.27 under the low wage situation and \$17.23 under the high wage situation if no taxes would need to be paid.

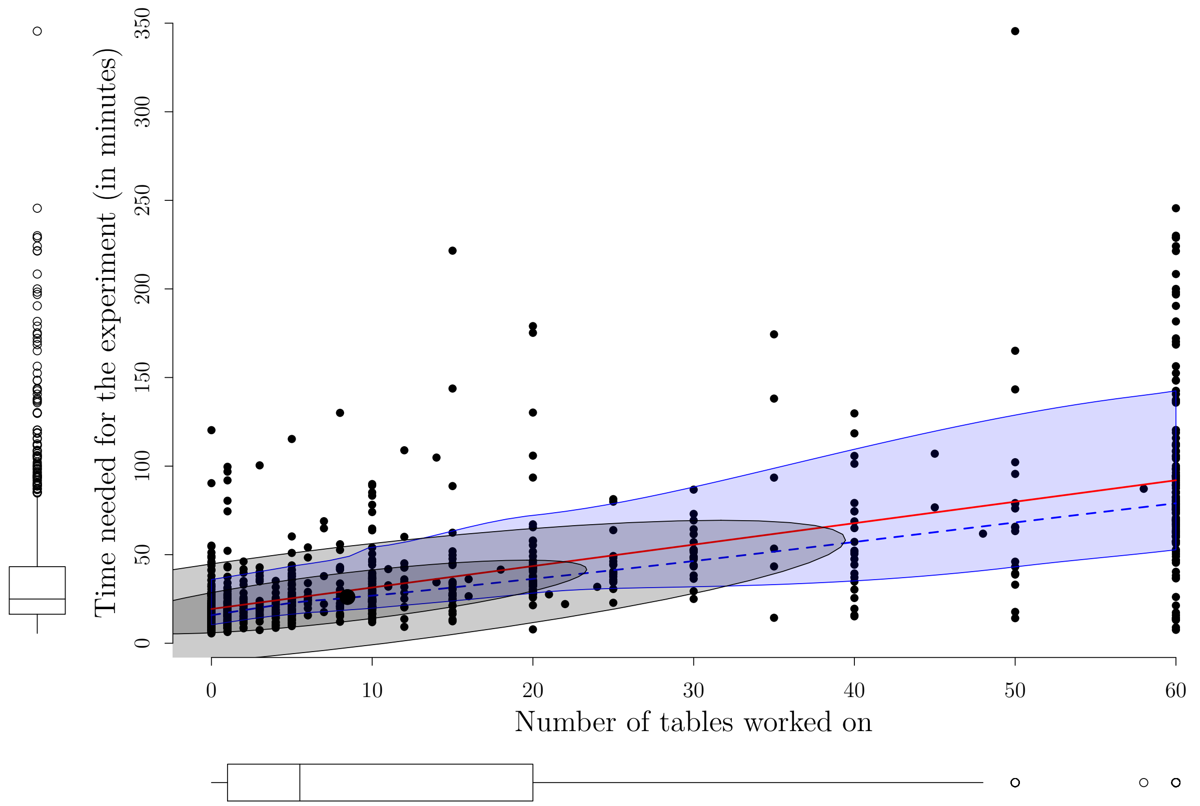
Comparing that to a target payment of about \$6 per hour for typical US-based MTurkers[43] we can see that high ability workers would be willing to work up to a tax of 65% under the high wage situation and up to a tax of 27% under the low wage situation. Low ability workers under the high wage situation would be willing to work up to a tax of 25% and under the low wage situation they would not be willing to work (as they would need to receive a subsidy of 56%). Thus, the incentives are structured such that there is sufficient room for labor supply responses (i.e. payoffs are better than typically for Mturkers) but also such that not all situations are worthwhile.

[43] See Berg (2015); Hornuf and Vrankar (2022).

A.3 Number of Worked Tables and Time Spent

An apparent divergence between our experiment and the underlying theory is that we assess labor supply based on the number of tasks completed, whereas the theory employs the duration of work. In this section, we demonstrate that these two measures are empirically equivalent, as there exists a robust positive correlation between the number of tasks solved and the time spent on them, as illustrated in Figure 8. Running a OLS reveals a strong and highly significant association (β=72.64,t(996)=28.02, $p \leq 0.001$, cohen's d=0.89) between the number of tables to be solved and the time needed to finish the experiment. Specifically, we see that participants needed on average 72.64 seconds longer to finish the experiment for each additional table worked.

Figure 8: Tables Worked and Time Spent.



Note: The figure shows the association between the number of tables worked on (on situation is randomly realized for participants and the corresponding decision is the number of tables assigned to work on). The x-axis denotes the number of tables worked on by participants, while the y-axis denotes the time needed overall to finish the whole experiment (in minutes). For each of the two measure the marginal boxplots are shown next to the respective axis. The black dotes in the main part of the figure denote individual observations. The red line depicts the linear regression line, while the blue line denotes the LOESS smoothing estimation with the corresponding confidence intervals. The gray bubbles denote data-concentration ellipses.

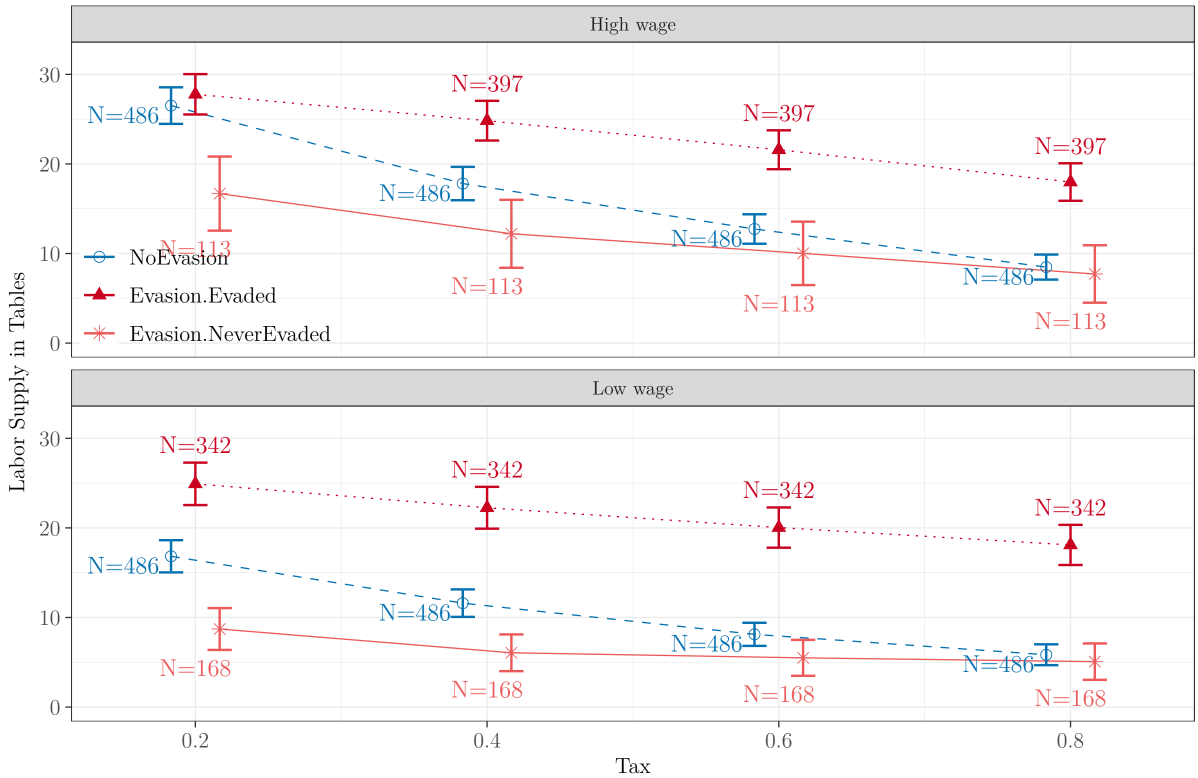
B Additional Analyses

B.1 Exploring Determinants of Self-Selection Into Evasion

For a deeper exploration of behavior relating to tax evasion, we compared the labor supply behavior of consistent non-evaders (i.e., those who never evaded in their wage level) and individuals who had evaded at least once in the Evasion-treatment. Our analysis reveals a notable divergence in labor supply patterns between these two distinct groups: Figure 9 depicts for both wage-levels the labor supply as a function the tax for these three situations. Subjects who never evaded had a substantially lower labor supply even for a tax of 20%. It is also evident that the behavior of those who decided not to evade and those who could not evade converged at a tax level of 60% and it is also evident that the labor supply of those who did evade was substantially higher compared to those who could not evade.

Building on these findings, we further examine the impact of various factors on evasion behavior and labor supply. Specifically, we concentrate on elements such as risk aversion, ability, personal income, and gender.

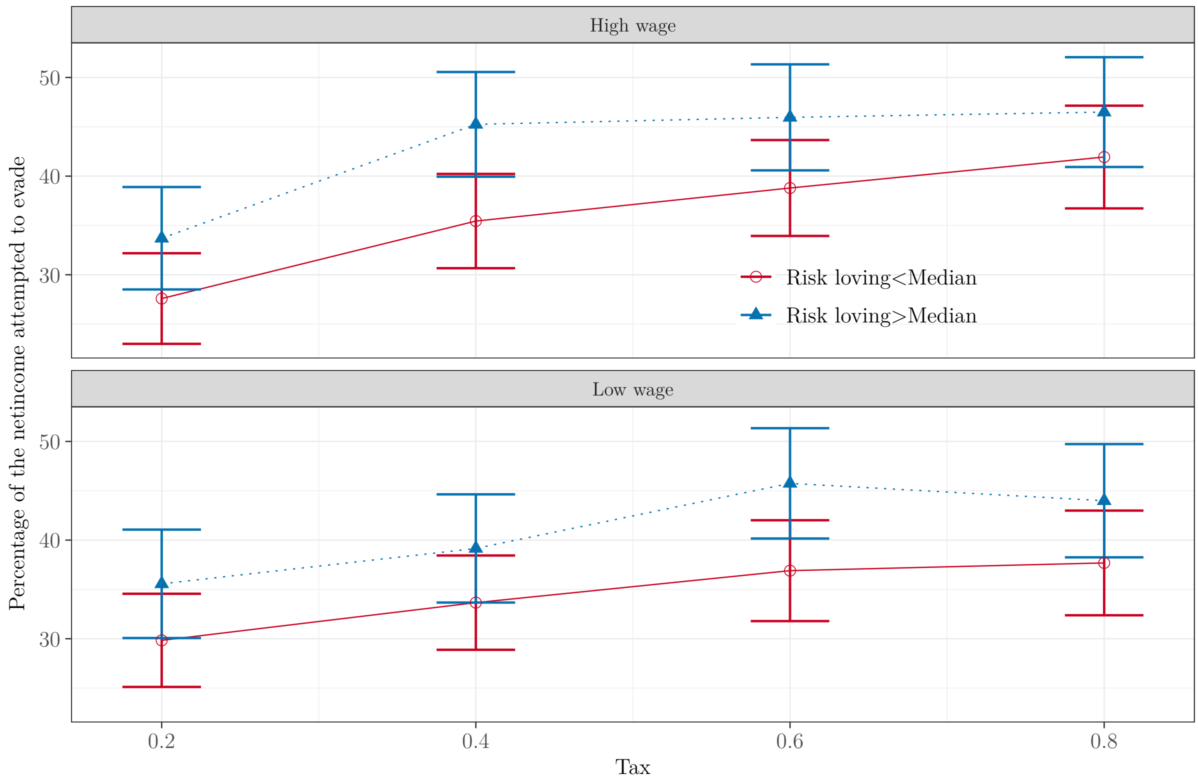
Figure 9: Labor Supply of Strict Non-Evaders.



Note: Orange, solid lines represent the Evasion-treatment where subjects never evaded (i.e. across all tax levels within a given wage level); blue, dashed lines represent the NoEvasion-treatment; red, dotted lines represent the Evasion-treatment where subjects evaded for at least one tax level. The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals.

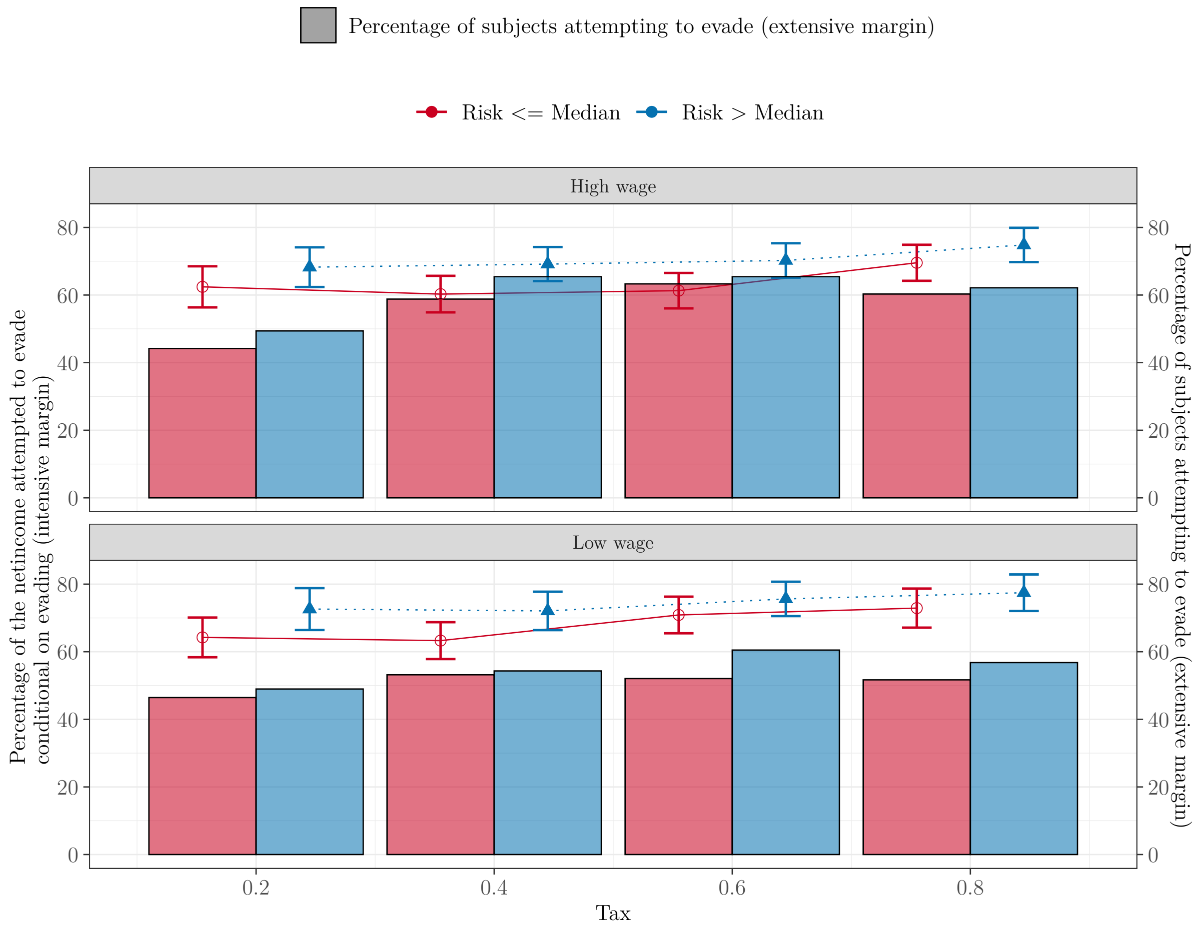
B.1.1 Risk Aversion

Figure 10: Tax Evasion by Risk Aversion (Median Split), Aggregated.



Note: Blue, dashed lines represent the percentage of the net income attempted to evade in the Evasion-treatment by subjects with risk measures above the median, while red, solid lines represent the percentage of the net income attempted to evade by subjects with risk measures below (or equal to) the median. The upper panel depicts the high wage situation while the panel at the bottom depicts the low wage situation. The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals.

Figure 11: Tax Evasion by Risk Aversion (Median Split), Conditional on Evading.



Note: Percentage of the net income attempted to evade in the Evasion-treatment conditional on evading as a function of tax with 95% confidence intervals in the Evasion-treatment. The lines in the bottom panel represent the percentage of the net income attempted to evade in the low wage situation, while the lines in the top panel represent the percentage of the net income attempted to evade for the high wage situation (intensive margin), all conditional on evading at all. The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals. The bars depict the percentage of subjects attempting to evade at the separate tax levels (i.e. the extensive margin). Blue colored elements represent subjects with risk measures below (or equal to) the median. Red colored elements represent subjects with risk measures above the median.

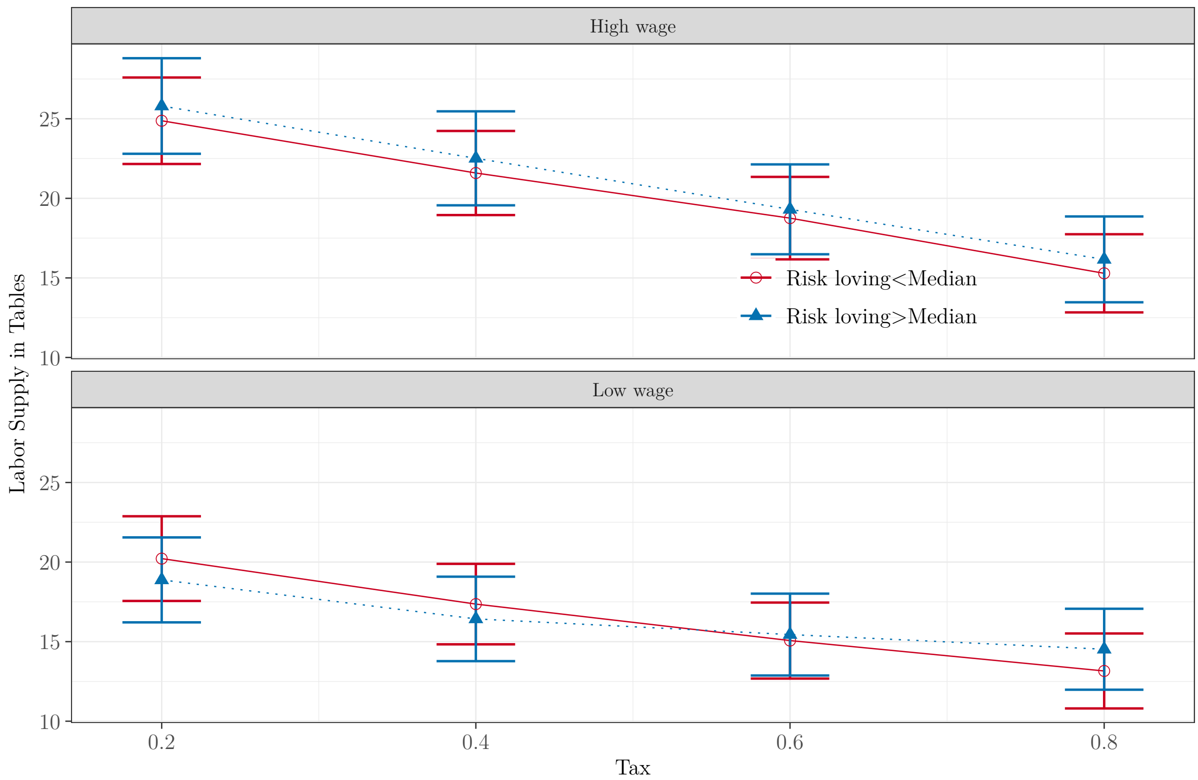
Table 6: Zero Inflated Normal Regression of Evasion Behavior as a Function of Risk.

	Extensive margin (i.e. Invest>0)								Intensive margin (i.e. Invest|Invest>0)							
	Low wage				High wage				Low wage				High wage			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Constant	−0.22	0.08	−0.03	−0.003	−0.36·	0.19	0.44*	0.35·	1.55***	1.42***	1.27***	1.09***	3.46***	2.07***	1.98***	1.90**
	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.21)	(0.20)	(0.34)	(0.32)	(0.31)	(0.31)	(0.72)	(0.62)	(0.59)	(0.63)
Risk	0.04	0.02	0.08	0.05	0.06	0.08·	0.04	0.03	0.11	0.13·	0.17*	0.20**	0.22	0.57***	0.51***	0.55***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.08)	(0.08)	(0.08)	(0.08)	(0.18)	(0.15)	(0.15)	(0.15)
Model	Log	Log	Log	Log	Log	Log	Log	Log	Norm	Norm	Norm	Norm	Norm	Norm	Norm	Norm
Tax	20%	40%	60%	80%	20%	40%	60%	80%	20%	40%	60%	80%	20%	40%	60%	80%
Observations	510	510	510	510	510	510	510	510	244	274	286	276	238	316	328	312
Adjusted R^2									0.003	0.01	0.01	0.02	0.002	0.04	0.03	0.04
Log Likelihood	−352.73	−352.00	−348.54	−351.32	−351.53	−337.38	−331.99	−340.49								

Notes: ·p<0.1;*p<0.05;**p<0.01;***p<0.001;

Risk denotes the elicited risk aversion with higher values indicating more risk-loving. Errors are clustered on the subject level, i.e., subject-specific effects do account for subject-specific heterogeneity.

Figure 12: Labor Supply by Risk Aversion (Median Split) and Wage Level.

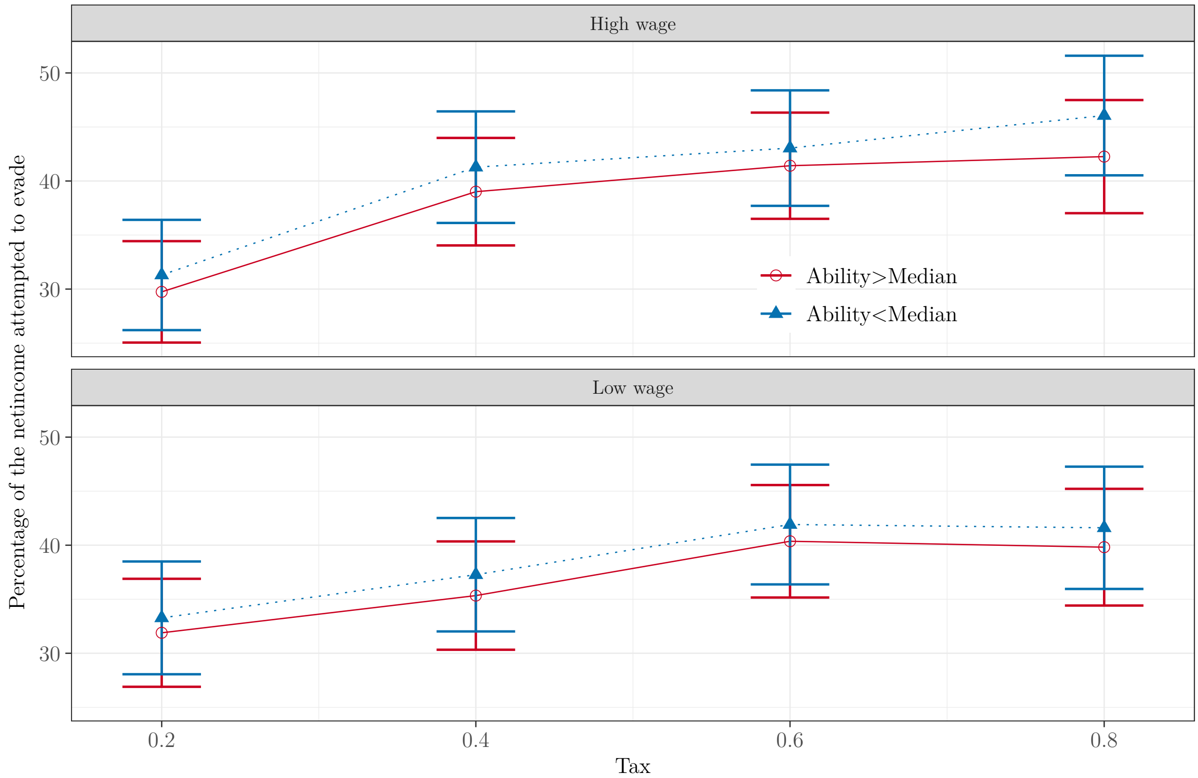


Note: Blue, dashed lines represent the labor supply by subjects with risk measures above the median, while red, solid lines represent the labor supply by subjects with risk measures below (or equal to) the median. The upper panel depicts the high wage situation while the panel at the bottom depicts the low wage situation. The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals.

B.1.2 Ability

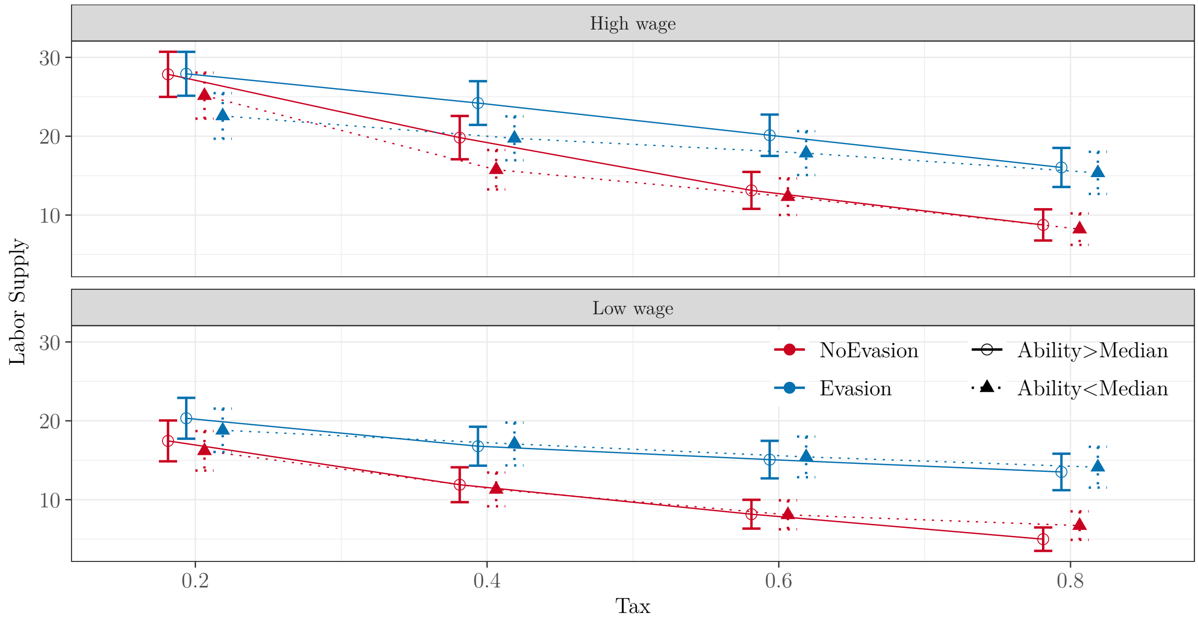
Are unproductive subjects are more inclined to engage in tax evasion? To explore this question, we regress on the decision to ever evade taxes (i.e. whether a given subject invested in any instance into the lottery) by the average time needed to solve the two trial tables. We find that the average time needed for the two trials has no effect on the decision to evade (β=-0.00,t(509)=-0.05, $p \geq 0.05$, cohen's d=0.00).

Figure 13: Evasion Decision by Ability.



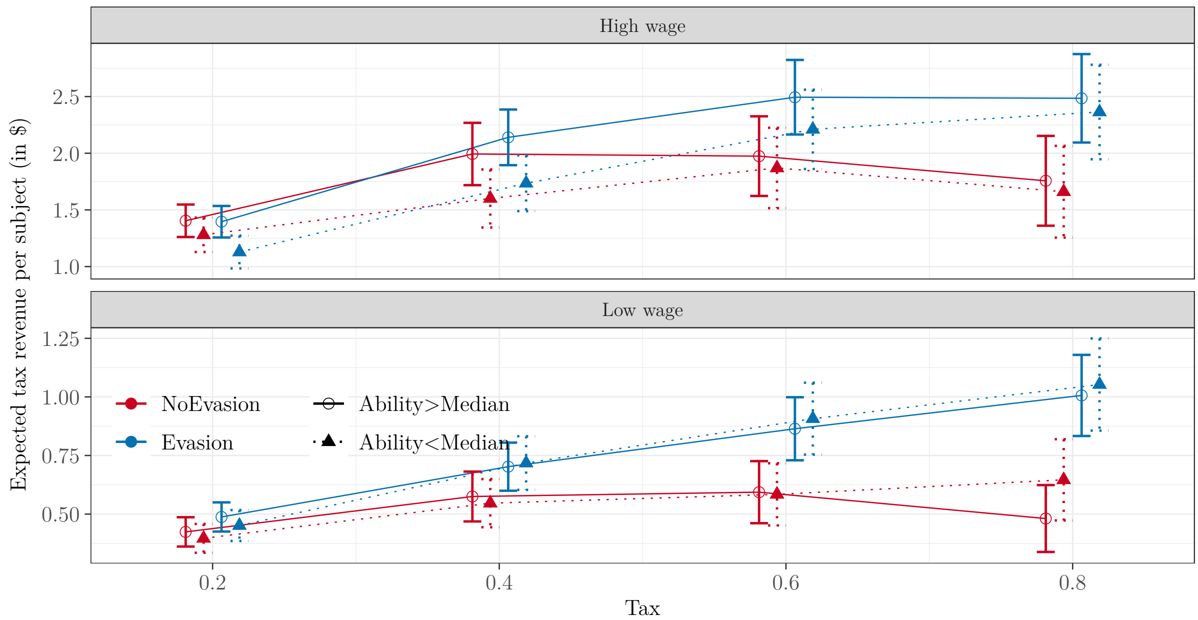
Note: Percentage of the net income attempted to evade as a function of tax in the Evasion-treatment for high and low ability subjects. Blue, dashed lines represent the percentage of the net income attempted to evade by subjects with a low ability of solving the tables (i.e. slower than the median subject), while red, solid lines represent the percentage of the net income attempted to evade by subjects with a high ability of solving the tables (i.e. faster than the median subject). The upper panel depicts the high wage situation while the panel at the bottom depicts the low wage situation. The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals.

Figure 14: Labor Supply by Ability.



Note: Blue lines represent the Evasion-treatment while red lines represent the NoEvasion-treatments. Triangles connected by dashed lines represent the labor supply of subjects with a low ability of solving the tables (i.e. slower than the median subject), while dots connected with solid lines represent the labor supply of subjects with a high ability of solving the tables (i.e. faster than the median subject). The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals. The upper panel depicts the high wage situation while the panel at the bottom depicts the low wage situation.

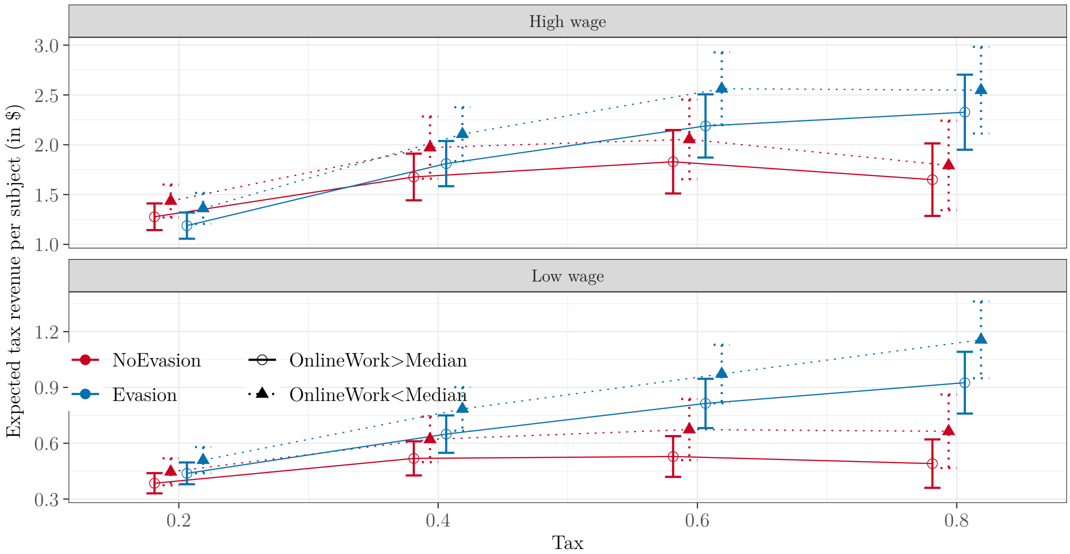
Figure 15: Expected Tax Revenue by Ability.



Note: Blue lines represent the Evasion-treatment while red lines represent the NoEvasion-treatments. Triangles connected by dashed lines represent the expected tax revenue of subjects with a low ability of solving the tables (i.e. slower than the median subject), while dots connected with solid lines represent the expected tax revenue of subjects with a high ability of solving the tables (i.e. faster than the median subject). The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals. The upper panel depicts the high wage situation while the panel at the bottom depicts the low wage situation.

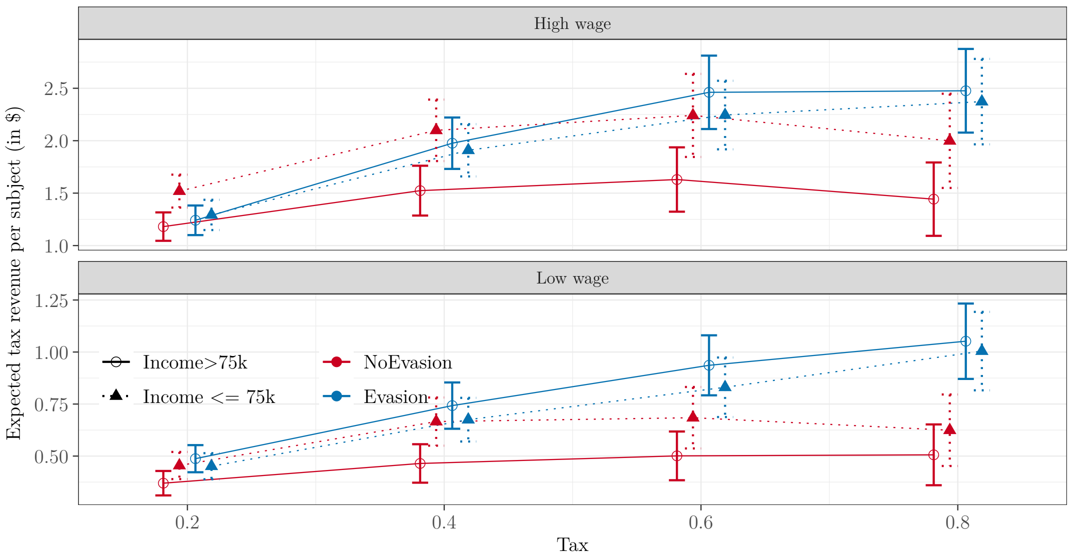
B.1.3 Personal Income Outside the Experimental Setting

Figure 16: Expected Tax Revenue by Hours Spent on Online Work.



Note: Expected tax revenue as a function of tax and treatment by hours spent on online work. Blue lines represent the Evasion-treatment while red lines represent the NoEvasion-treatments. Triangles connected by dashed lines represent the expected tax revenue of subjects with a few hours of online works (i.e. fewer than the median subject), while dots connected with solid lines represent the expected tax revenue of subjects with a lot of hours of online works (i.e. higher than the median subject). The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals. The upper panel depicts the high wage situation while the panel at the bottom depicts the low wage situation.

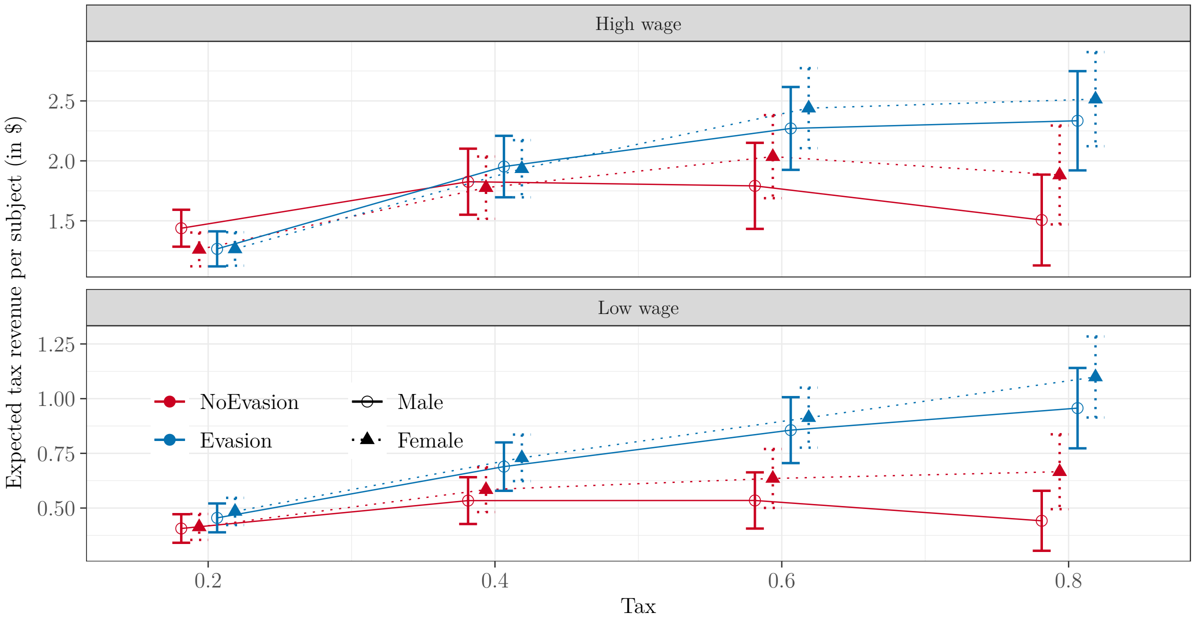
Figure 17: Expected Tax Revenue by Income.



Note: Expected tax revenue as a function of tax and treatment by income. Blue lines represent the Evasion-treatment while red lines represent the NoEvasion-treatments. Triangles connected by dashed lines represent the expected tax revenue of subjects with a lower income (i.e. annual income of less than 75k), while dots connected with solid lines represent the expected tax revenue of subjects with a higher income (i.e. annual income of more than 75k). The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals. The upper panel depicts the high wage situation while the panel at the bottom depicts the low wage situation.

B.1.4 Gender

Figure 18: Expected Tax Revenue by Gender.



Note: Expected tax revenue as a function of tax and treatment by gender. Blue lines represent the Evasion-treatment while red lines represent the NoEvasion-treatments. Triangles connected by dashed lines represent the expected tax revenue of female subjects, while dots connected with solid lines represent the expected tax revenue of male subjects. The corresponding tunnels surrounding the respective dots represent the 95% confidence intervals. The upper panel depicts the high wage situation while the panel at the bottom depicts the low wage situation.

B.2 Tax Revenue - Non-Linear Regression Models

Table 7: Cubic Mixed-Effects Model of Expected Tax Revenue.

	Expected Tax revenue per subject (in $)					
		Low-Wage			High-Wage	
Constant	0.53*** (0.04)	0.44*** (0.09)	0.58*** (0.17)	1.69*** (0.10)	1.29*** (0.20)	2.06*** (0.41)
as.factor(Treatment2)Evasion	0.24*** (0.06)	0.29* (0.12)	0.27* (0.12)	0.30* (0.13)	0.53· (0.29)	0.51· (0.29)
poly(Tax2, 2, raw = FALSE)1	3.42*** (0.81)	−0.56 (1.74)	−0.56 (1.74)	8.62*** (2.03)	4.91 (4.35)	4.91 (4.35)
poly(Tax2, 2, raw = FALSE)2	−2.79*** (0.81)	−3.27· (1.74)	−3.27· (1.74)	−10.57*** (2.03)	−9.85* (4.35)	−9.85* (4.35)
as.factor(Treatment2)Evasion:poly(Tax2, 2, raw = FALSE)2	1.28 (1.14)	0.95 (2.49)	0.95 (2.49)	0.98 (2.84)	−0.61 (6.22)	−0.61 (6.22)
as.factor(Treatment2)Evasion:poly(Tax2, 2, raw = FALSE)1	9.65*** (1.14)	11.87*** (2.49)	11.87*** (2.49)	18.85*** (2.84)	21.31*** (6.22)	21.31*** (6.22)
as.factor(Treatment2)Evasion:Risk		−0.02 (0.03)	−0.01 (0.03)		−0.07 (0.08)	−0.07 (0.08)
Risk		0.03 (0.02)	0.03 (0.02)		0.12* (0.05)	0.11* (0.05)
poly(Tax2, 2, raw = FALSE)1:Risk		1.22** (0.47)	1.22** (0.47)		1.13 (1.18)	1.13 (1.18)
poly(Tax2, 2, raw = FALSE)2:Risk		0.15 (0.47)	0.15 (0.47)		−0.22 (1.18)	−0.22 (1.18)
as.factor(Treatment2)Evasion:poly(Tax2, 2, raw = FALSE)1:Risk		−0.73 (0.64)	−0.73 (0.64)		−0.79 (1.61)	−0.79 (1.61)
as.factor(Treatment2)Evasion:poly(Tax2, 2, raw = FALSE)2:Risk		0.08 (0.64)	0.08 (0.64)		0.46 (1.61)	0.46 (1.61)
Controls	×	×	✓	×	×	✓
Sbj specific effects	✓	✓	✓	✓	✓	✓
Observations	3,984	3,984	3,984	3,984	3,984	3,984
Log Likelihood	−4,536.59	−4,536.76	−4,555.62	−8,119.52	−8,115.88	−8,126.36
Akaike Inf. Crit.	9,089.18	9,101.52	9,157.24	16,255.04	16,259.76	16,298.72
Bayesian Inf. Crit.	9,139.50	9,189.58	9,301.91	16,305.36	16,347.83	16,443.39

Notes: ·$p < 0.1$;*$p < 0.05$;**$p < 0.01$;***$p < 0.001$;

Note: Here we use an orthogonal polynomial regression with degree two. Controls include age, gender, ethnicity, income, party affiliation, employment status, education, hours spent on online work, and the average time needed for solving the two sample tasks. Evasion denotes a dummy with value one if the participant was in the Evasion-treatment – participants have the opportunity to evade taxes and punishment is implemented with a 50% probability – and zero otherwise. The omitted category are the NoEvasion-treatments. *Tax* denotes a one percentage point increase in the tax level. *Risk* denotes the elicited risk aversion with higher values indicating more risk-loving. Errors are clustered on the subject level, i.e., subject-specific effects do account for subject-specific heterogeneity.

Table 8: GAM Mixed-Effects Model of Expected Tax Revenue.

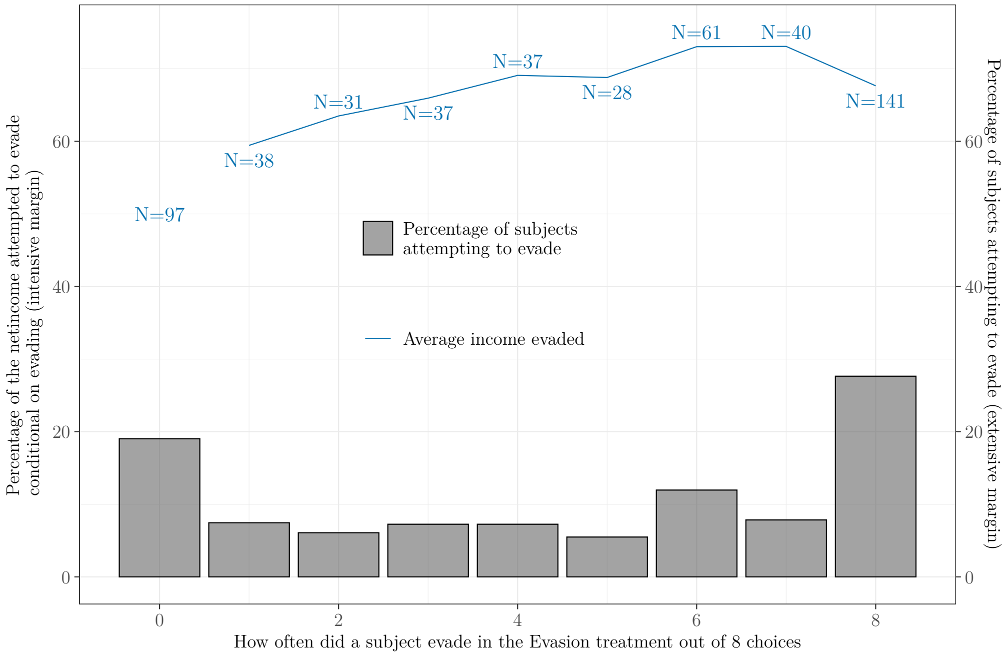
	Expected Tax revenue per subject (in $)					
		Low-Wage			High-Wage	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.53**** (0.04)	0.46**** (0.07)	0.60**** (0.16)	1.69**** (0.09)	1.42**** (0.15)	2.17**** (0.39)
Evasion	0.24**** (0.06)	0.24**** (0.06)	0.24**** (0.06)	0.30** (0.13)	0.28** (0.13)	0.28** (0.13)
Risk		0.02 (0.02)	0.02 (0.02)		0.08** (0.04)	0.07** (0.04)
Controls	×	×	✓	×	×	✓
Sbj specific effects	✓	✓	✓	✓	✓	✓
Observations	3984	3984	3984	3984	3984	3984
Log Likelihood	-4574.15	-4573.34	-4563.31	-8149.53	-8147.08	-8136.33

Notes: ·$p < 0.1$;*$p < 0.05$;**$p< 0.01$;***$p < 0.001$;

The table depicts the estimation results of a generalized additive model (GAM) on the effect of taxes on tax revenues. Thin plate regression splines are used to account for the non-linear function of the slope of the tax revenues. Controls include age, gender, ethnicity, income, party affiliation, employment status, education, hours spent on online work, and the average time needed for solving the two sample tasks.Evasion denotes a dummy with value one if the participant was in the Evasion-treatment – participants have the opportunity to evade taxes and punishment is implemented with a 50% probability – and zero otherwise. The omitted category are the NoEvasion-treatments. *Risk* denotes the elicited risk aversion with higher values indicating more risk-loving. Errors are clustered on the subject level, i.e., subject-specific effects do account for subject-specific heterogeneity.

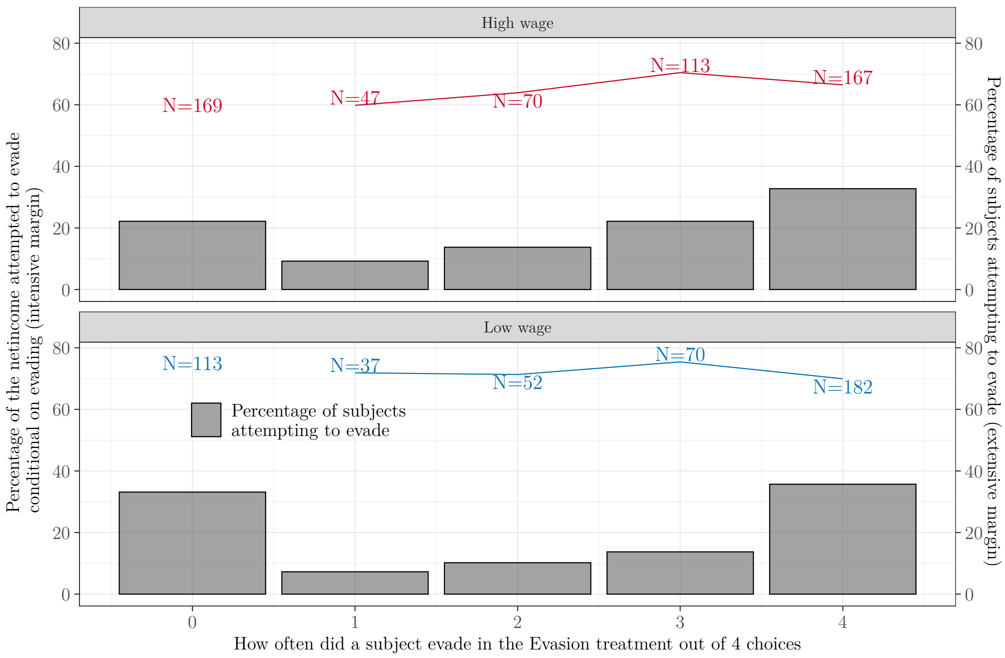
B.3 Within-Subject Distribution of the Evasion Decision

Figure 19: Individual Evasion Decisions; Extensive and Intensive Margin.



Note: Blue, solid lines represent the percentage of the net income attempted to evade conditional on evading at all. The corresponding numbers indicate the number of people evading one, two, etc. times. The gray bars depict the percentage of subjects attempting to evade one, two, etc. times (i.e. the extensive margin).

Figure 20: Individual Evasion Decisions by Wage Level.

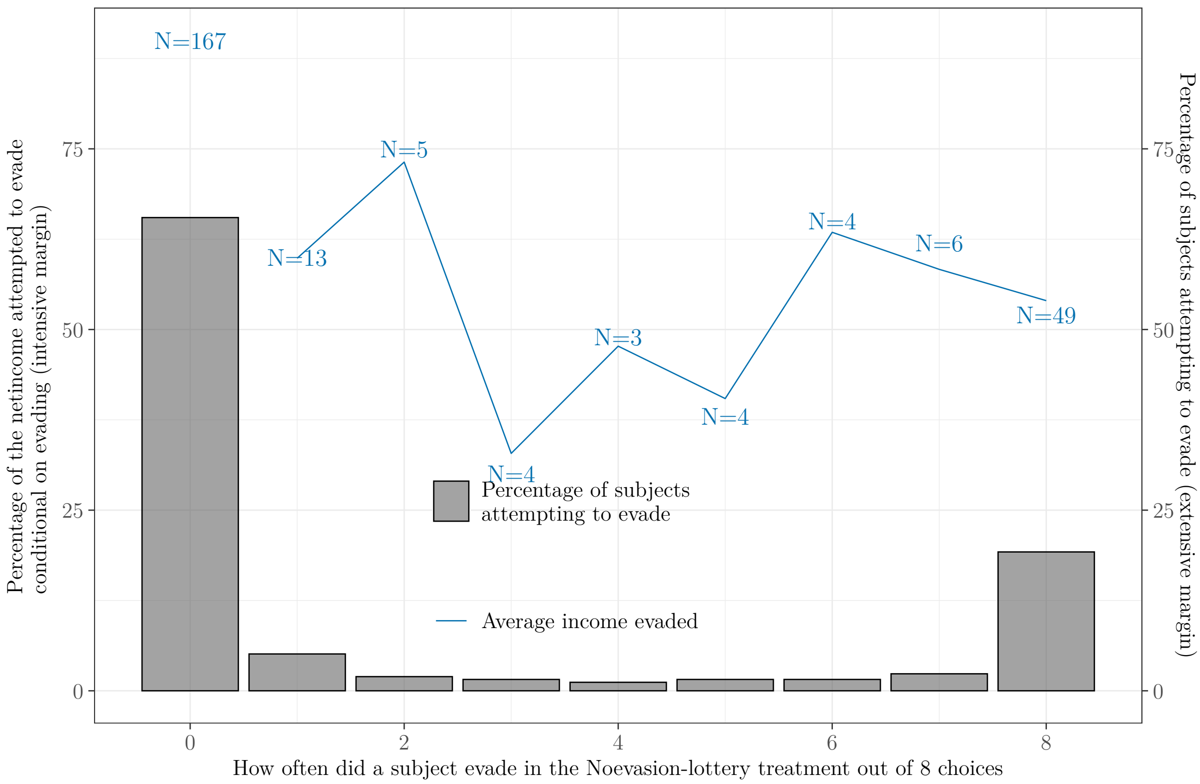


Note: Blue/red, solid lines represent the percentage of the net income attempted to evade conditional on evading at all in the high/low wage situation. The corresponding numbers indicate the number of people evading one, two, etc. times. The gray bars depict the percentage of subjects attempting to evade one, two, etc. times (i.e. the extensive margin).

C Difference in the Two No-Evasion Treatments

The NoEvasion-Lottery-treatment was purely designed to ensure that the behavior is not driven by longer instructions in the Evasion-treatment, the availability of another option in the Evasion-treatment or the reduced cognitive load in the NoEvasion-treatments.[44] In fact, two NoEvasion-treatments do not differ significantly from each other: The average labor supply was $M = 14.25$ ($SD = 16.15$) in the NoEvasion-Lottery-treatment and $M = 12.94$ ($SD = 14.39$) in the NoEvasion-NoLottery-treatment, $t(411.3)= 0.9$, $p{\geq}0.05$. They do not differ in any of the eight possible situations. All the results also go through if we just compare the Evasion-treatment to the main evasion treatment, i.e. NoEvasion-NoLottery.

Figure 21: Individual Evasion Decisions in the NoEvasion-Lottery-Treatment; Extensive and Intensive Margin.



Note: Percentage of the net income attempted to evade conditional on evading as a function of how often a subject evaded. Blue, solid lines represent the percentage of the net income attempted to evade conditional on evading at all. The corresponding numbers indicate the number of people evading one, two, etc. times. The gray bars depict the percentage of subjects attempting to evade one, two, etc. times (i.e. the extensive margin).

[44] Abeler and Jäger (2015) show how the complexity in tax system might change the reaction function to changes in tax rates.

D Experimental Design

D.1 Experimental Procedure

This section describes the chronological order of the actual experiment in more detail. The sequence of the specific experimental procedure is depicted in Figure 22 in the Appendix.

Preceding Task Description. Before the start of our study, participants were informed about the general structure of the experiment and their expected payoff with the following information: The experiment consists of two stages, the first lasting about 15 minutes comprising some basic demographic questions, a short "mini experiment" (i.e. a small game to elicit the risk aversion) and a decision on how long to work in the second stage. In the second stage they would simply have to perform the amount of work indicated in the first stage, which could be between 0 and 45 additional minutes. For the first stage the participants received an average compensation of $2.50, whereas the payment of the second stage was conditional on the amount of work they choose to perform.

Demographics. At the beginning of this first stage, our participants had to answer a short questionnaire on standard socio-economic background variables. These are: gender, age, education, employment status, household income, state, ethnicity, political orientation. Moreover, we asked for the average hours per week they spend doing online tasks for money in order to assess how familiar they are with such microtasks.

Risk Elicitation. Subsequently, the risk preferences of the participants were elicited using a version of the Eckel and Grossman (2002) method suggested by Dave et al. (2010).[45] Participants were faced with a choice set of six different gambles with each gamble involving a 50% chance of winning an either high or low payment. Only the first gamble was a safe bet with a $1.65 payoff. The following choice options increased linearly in expected payoff but also in risk, with an expected payoff of $2.25 in the riskiest scenario ($0 or $4.50). In their overview article Charness et al. (2013) describe this method as relatively easy to understand and thus producing less noisy estimates of risk preferences than other elicitation methods, especially when participants have low math abilities. A relevant advantage given the more diverse non-student sample of MTurk.

Trail Stage. Before having continued to the actual work decisions, our participants were asked to count the zeros of two tables in order to familiarize themselves with the upcoming task. In this trial period we measured the individual time needed to proceed to the following table in order to give individual feedback on the average time required per table but also to be able to account for "ability" in our analysis.

Work Decisions and Work Stage. As described above, participants took 8 work decisions, of which only one was randomly selected actually had to be performed. Apply-

[45]Importantly, we counterbalanced if this risk elicitation decision had to be made before or after the work and evasion decisions took place in order to control for potential risk-hedging strategies between our participants. Our data indicated no order effect on the risk preferences nor any income effect (i.e. no effect of the order on the labor supply).

ing this strategy method allows to elicit multiple informative data points per subject (i.e. decisions for eight different payment schemes), without confounding effects like exhaustion due to working on all eight payment schemes. It is important to note that our instructions strongly emphasized how the payment would only be made when all the indicated tables were counted correctly (i.e. only then the MTurk code which qualified the payment was displayed). To help gauge the individual time required, participants received individual feedback on the estimated duration based on their average time per table in the preceding trial period.[46]

Final Payoff. Feedback on the final payoff was only provided at the very end of our experiment to cancel out wealth effects that may have distorted work supply decisions and risk aversion. The final payment consisted of a $0.50 fixed payment, the outcome of the risk-elicitation lottery (which amounted to an average compensation of $2.50 together with the former) as well as the outcome of the work/evasion decision. Thus, the possible final payments ranged between $0.50 and $20.00 if the participants picked the riskiest decisions throughout.

In the following section, specific instructions are provided verbatim.

D.2 Instructions Trial Stage

Below you see a 10x15 table with zeros and ones. We would like you to count the numbers of zeros. **Only if you entered the right amount, you will be able to proceed to the next page**.
Please count the numbers of zeros for two subsequent tables on the next pages. This will familiarize you with the task in order to indicate how many of these tables you want to work on in the second part of this HIT. You will be paid per table in the second part of our study.

D.3 Instructions Evasion-Treatment

Good job! It took you on average [**seconds needed in trial**] seconds per table.

Now, we would like to know how many of these tables you would like to work on in the second part of this study.
To do so, we want you to indicate your preferred number of tables for **eight different payment schemes. After taking your eight decisions, you will have to work on your decision for only one randomly picked payment scheme!**
These payment schemes differ in:

- The payment per table and

[46] We set a limit to a max. of 60 tables. Based on a pilot study, the experiment then takes around 60 minutes in total.

- A proportionate fee you have to pay to the requester - i.e., how much of your earned money per table you can keep.

You can also avoid the payment of the fee. To do so, you can participate in a **lottery**. The lottery is represented by a fair coin toss – you have a 50% chance to win, and you have a 50% chance to lose. On any amount **not invested** in the lottery you will need to pay the fee. For any amount **invested** in the lottery the following holds:

- **If you win:** you avoid paying any fee on the invested amount.
- **If you lose:** you will need to pay the fee on the invested amount plus an additional fee of 20% on the money invested.

Thus, you will need to make two decisions in this part of the study:

1. How many tables you would like to solve (which will result in a gross payment).
2. How much money of the gross payment you want to invest in a fair lottery (i.e on which part of your income you would like to attempt to avoid the fees).

You can freely decide on how many tables you would like to work on, with at most 60 tables. Thus, you can also decide not to work at all and correspondingly you would obtain only the payment for the mini-experiment and the $0.50 fixed payment.
As already mentioned, you will make the two decisions under **8 different payment schemes**. The payment schemes differ in the **payment per table** and the **proportionate fee**.
The payment per table will be either $0.12 or $0.25 for every correctly solved table. Each of these two levels is shown in a block of 4 scenarios (e.g. Block I: $0.12 and Block II: $0.25). Thus, you can earn between additional $0 (if you decide to work on 0 tables) and maximally $15 (if you decide on 60 tables) gross.
The **proportionate fee** will be either 20%, 40%, 60% or 80%. Each of these four levels will be shown in both blocks (= eight scenarios). The proportionate fee indicates how much of your gross earnings you will be effectively paid by the end of the experiment.
For example: if the proportionate fee is 60% you will obtain only 40% of your gross income. So if you earned $5 gross, you would obtain only $5*40/100=$2 by the end of the experiment.
Strategic Advice:
Given the odds of the lottery (50% / 50%) and the additional fee of 20% if the lottery is lost, it might not be profitable to invest in the lottery if the regular fee is low (20%). However, with higher regular fees (40%, 60% or 80%) investments into the lottery might be profitable.
At the end of the experiment only ONE of the 8 different payment schemes will be made payoff-relevant for you. You will be informed about which one is payoff-relevant before working on the tables.

D.4 Instructions NoEvasion-Lottery-Treatment

Good job! It took you on average [**seconds needed in trial**] seconds per table.

Now, we would like to know how many of these tables you would like to work on in the second part of this study.
To do so, we want you to indicate your preferred number of tables for **eight different payment schemes. After taking your eight decisions, you will have to work on your decision for only one randomly picked payment scheme!**
These payment schemes differ in:

- The **payment per table** and
- A proportionate **fee** you have to pay to the requester - i.e., how much of your earned money per table you can keep.

You can participate in a lottery to avoid the fees. **However, you have a 100% chance to lose**. On any amount **not invested** in the lottery you will need to pay the fee. For any amount **invested** in the lottery you will need to pay the fee on the invested amount plus an additional fee of 20% on the money invested.
Thus, you will need to make two decisions in this part of the study:

1. How many tables you would like to solve (which will result in a gross payment).
2. How much money of the gross payment you want to invest in **the lottery you will always lose**.

You can freely decide on how many tables you would like to work on, with at most 60 tables. Thus, you can also decide not to work at all and correspondingly you would obtain only the payment for the mini-experiment and the $0.50 fixed payment.
As already mentioned you will make the two decisions under **8 different payment schemes**. The payment schemes differ in the **payment per table** and the **proportionate fee**.
The payment per table will be either $0.12 or $0.25 for every correctly solved table. Each of these two levels is shown in a block of 4 scenarios (e.g. Block I: $0.12 and Block II: $0.25). Thus, you can earn between additional $0 (if you decide to work on 0 tables) and maximally $15 (if you decide on 60 tables) gross.
The **proportionate fee** will be either 20%, 40%, 60% or 80%. Each of these four levels will be shown in both blocks (= eight scenarios). The proportionate fee indicates how much of your gross earnings you will be effectively paid by the end of the experiment.
For example: if the proportionate fee is 60% you will obtain only 40% of your gross income. So if you earned $5 gross, you would obtain only $5*40/100=$2 by the end of the experiment.
Strategic Advice:
Given the odds of the lottery (i.e. you will never win) and the additional fee of 20%, it is

not profitable to invest into the lottery because you would definitely lose money.
At the end of the experiment only ONE of the 8 different payment schemes will be made payoff-relevant for you. You will be informed about which one is payoff-relevant before working on the tables.

D.5 Instructions NoEvasion-NoLottery-Treatment

Good job! It took you on average [**seconds needed in trial**] seconds per table.

Now, we would like to know how many of these tables you would like to work on in the second part of this study.
To do so, we want you to indicate your preferred number of tables for **eight different payment schemes. After taking your eight decisions, you will have to work on your decision for only one randomly picked payment scheme!**
These payment schemes differ in:

- The payment per table and
- A proportionate fee you have to pay to the requester - i.e., how much of your earned money per table you can keep.

You can freely decide on how many tables you would like to work on, with at most 60 tables. Thus, you can also decide not to work at all and correspondingly you would obtain only the payment for the mini-experiment and the $0.50 fixed payment.
As already mentioned you will make the two decisions under **8 different payment schemes**. The payment schemes differ in the **payment per table** and the **proportionate fee**.
The payment per table will be either $0.12 or $0.25 for every correctly solved table. Each of these two levels is shown in a block of 4 scenarios (e.g. Block I: $0.12 and Block II: $0.25). Thus, you can earn between additional $0 (if you decide to work on 0 tables) and maximally $15 (if you decide on 60 tables) gross.
The **proportionate fee** will be either 20%, 40%, 60% or 80%. Each of these four levels will be shown in both blocks (= eight scenarios). The proportionate fee indicates how much of your gross earnings you will be effectively paid by the end of the experiment.
For example: if the proportionate fee is 60% you will obtain only 40% of your gross income. So if you earned $5 gross, you would obtain only $5*40/100=$2 by the end of the experiment.
At the end of the experiment only ONE of the 8 different payment schemes will be made payoff-relevant for you. You will be informed about which one is payoff-relevant before working on the tables.

D.6 Experimental Procedure

Figure 22: Sequence of the Experimental Procedure.

Socioeconomic Background Questions

⇓

Eliciting Risk Preferences

Eckel and Grossmann Measure

⇓

Instructions I: Testing 2 Tables

⇓

Instructions II: Piece-Rates, Fees & Lottery

Between Subject Design: Perfect vs. Imperfect "Tax Enforcement"

⇓

Comprehension Control Questions

⇓

Work Decisions

8 Decisions: 4 Low & 4 High Piece-Rate

⇓

Work Stage

⇓

Feedback on Final Payoff

Risk Elicitation & Income Lottery

D.7 Screenshots

Figure 23: Screenshot of the Real Effort Task.

Table not displayed?

Reload Page

Number of tables: 1 / 2

Please count the number of zeros and enter your result in the box below:

0	0	1	1	0	1	1	1	0	0	0	1	1	1	1
0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
0	0	1	0	0	0	1	0	1	0	0	1	1	1	1
0	1	1	0	1	0	0	0	1	0	1	0	1	1	1
0	0	1	0	0	0	1	0	1	1	0	1	1	1	0
1	1	1	1	1	0	0	1	1	1	1	0	0	1	0
0	1	1	1	1	1	1	0	0	0	0	0	1	1	0
0	0	1	0	0	1	0	0	1	1	0	1	0	0	0
1	0	0	0	0	1	0	0	0	1	0	1	0	0	1
0	1	0	1	1	0	0	0	0	0	0	0	1	1	1

→

Figure 24: Screenshot of the Evasion-Treatment.

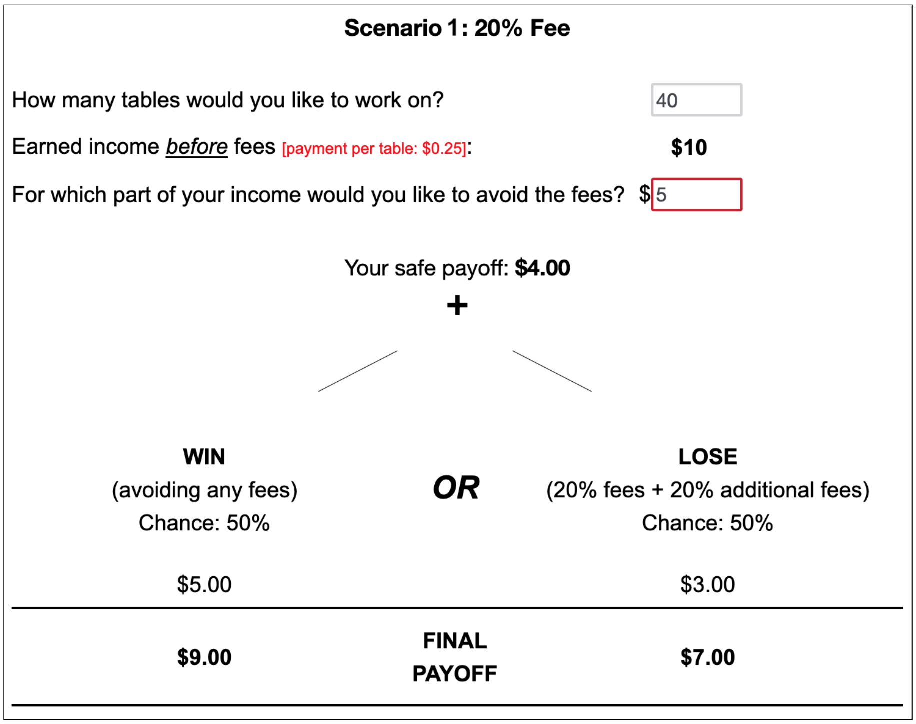


Figure 25: Screenshot of the NoEvasion-Lottery-Treatment.