Ca' Foscari
University
of Venice

Department
of Economics

Working Paper

Michele Bernasconi
Irene Maria Buso
Anna Marenzi
Dino Rizzi

Tax Notches in the Lab: Disentangling Real and Evasion Responses

ISSN: 1827-3580
No. 15/WP/2025

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 15/WP/2025
ISSN 1827-3580

Tax Notches in the Lab: Disentangling Real and Evasion Responses

Michele Bernasconi
Ca' Foscari University of Venice

Irene Maria Buso
University of Bologna

Anna Marenzi
Ca' Foscari University of Venice

Dino Rizzi
Ca' Foscari University of Venice

Abstract
Several empirical studies find large behavioural responses to the incentives created by tax notches, highlighting the challenge of distinguishing between responses driven by real effects and by tax reporting. In a lab experiment, we find strong evidence of excessive bunching in a tax notch system, both with and without evasion possibilities. The effort adjustments are mainly from above the threshold, while the evasion adjustments are mainly from below the threshold. Both adjustments contribute to a reduction in the underreporting rate. We also confirm evidence of optimisation frictions. They are stronger when evasion is possible than when it is not. A gender breakdown highlights both the robustness of the effects found and the impact that heterogeneous preferences can have on the overall responses to tax notches.

Keywords
Tax Evasion, Effort, Bunching, Lab Experiment

JEL Codes
H24, H26, C91

Address for correspondence:
Michele Bernasconi
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
e-mail: bernasconi@unive.it

The Working Paper Series
is available only on line
(https://www.unive.it/dec-workingpapers)
For editorial correspondence, please contact:
wp.dse@unive.it

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta San Giobbe
30121 Venice Italy
Fax: ++39 041 2349210

Tax Notches in the Lab: Disentangling Real and Evasion Responses

Michele Bernasconi[a] Irene Maria Buso[b] Anna Marenzi [*] Dino Rizzi [§]

August 21, 2025

Abstract

Several empirical studies find large behavioural responses to the incentives created by tax notches, highlighting the challenge of distinguishing between responses driven by real effects and by tax reporting. In a lab experiment, we find strong evidence of excessive bunching in a tax notch system, both with and without evasion possibilities. The effort adjustments are mainly from above the threshold, while the evasion adjustments are mainly from below the threshold. Both adjustments contribute to a reduction in the underreporting rate. We also confirm evidence of optimisation frictions. They are stronger when evasion is possible than when it is not. A gender breakdown highlights both the robustness of the effects found and the impact that heterogeneous preferences can have on the overall responses to tax notches.

Keywords: Tax Evasion, Effort, Bunching, Lab Experiment
JEL Classification: H24, H26, C91

[a]Department of Economics, Ca' Foscari University of Venice, San Giobbe, Cannaregio 873, 30121 Venezia, e-mail: bernasconi@unive.it.

[b]Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna, 40126, e-mail: irenemaria.buso@unibo.it

[*]Department of Economics, Ca' Foscari University of Venice, San Giobbe, Cannaregio 873, 30121 Venezia, e-mail: anna.marenzi@unive.it.

[§]Department of Economics, Ca' Foscari University of Venice, San Giobbe, Cannaregio 873, 30121 Venezia, e-mail: rizzid@unive.it.

1 Introduction

An area of tax policy recently attracting renewed interest concerns the effects of systems with discontinuous jumps in taxpayer liability, known as 'tax notches' (Blinder and Rosen, 1985; Kleven and Waseem, 2013; Slemrod, 2013). These jumps typically create an area of dominated decisions, giving taxpayers an incentive to move towards the more attractive side of the notch. Despite the distortions this entails in a world where first-best policies could be implemented, many tax systems worldwide include notches. While in some cases such notches are unintended or incidental, in others, they can have justifications (Slemrod, 2013; Kanbur and Keen, 2014; Kleven, 2016).

A case concerns income tax regimes, which target micro and small businesses by offering a preferential low flat rate for income up to a certain threshold, with the standard regime applying if income exceeds this threshold. Justifications for these schemes are based on simplicity and cost reduction, often accompanied by the expectation of the tax authorities who choose these policies that a lower level of taxation, made salient by a discrete, visible drop in the tax rate, could encourage tax compliance and even facilitate the transition of 'hard-to-tax' individuals from the informal to the formal economy. Traditionally recommended for developing countries, income tax systems with these regimes are increasingly adopted by advanced economies as well (e.g., Austria, France, Italy, Portugal; OECD, 2023).[1]

However, these systems may also create the opposite incentive, to falsely report at the tax threshold to benefit from the preferential regime. As a result, taxpayers' bunching at the tax threshold could originate from both real and evasion adjustments, as well as from adjustments that increase or decrease total tax evasion.

Unfortunately, theoretical analysis hardly helps to clarify bunching's origin in this context. In classical models of tax evasion behaviour (Allingham and Sandmo, 1972; Pencavel, 1979), for instance, substitution and income effects on labour supply and reporting decisions lead to ambiguous responses. Similarly, the theoretical framework developed in Section 2 shows that, even when taxpayers' behaviour is restricted by a quasi-linear preference model commonly used in empirical research, the theoretical responses to the tax notch remain indeterminate, and the very same area of dominated

[1] These schemes are sometimes also called presumptive income tax regimes since notch thresholds are expressed in terms of indices, such as business turnovers, from which income is determined according to presumptive coefficients. Nevertheless, the decisions in the schemes on how much turnover to report and whether to report it above or below the threshold remain with taxpayers. (These schemes also differ from tax systems in which taxpayers are taxed on either profits or turnover depending on which tax liability is larger. Such systems, which generate kinks rather than notches, can also lead to a reduction of evasion; Best et al., 2015).

bundles becomes less universal when tax evasion is possible.

Empirically, extensive recent literature confirms bunching at notch thresholds while facing the challenge of disentangling its origin due to the difficulty of observing tax evasion with field data. Kleven and Waseem (2013) provided the seminal paper in this area, developing an approach to estimate structural elasticities from observed bunching at personal income tax notch thresholds in Pakistan.[2] They observed significant bunching, despite small elasticities and considerable frictions. Adapting the methodology, large notch responses have been consistently found across various tax systems and different taxes (Best and Kleven, 2018; Liu et al., 2021; Adam et al., 2021; Garbinti et al., 2023; Muthitacharoen et al., 2021; Tazhitdinova, 2020; Bachas and Soto, 2021; Lobel et al., 2024). While, without tax evasion, bunching in the approach can be entirely interpreted as a real response, studies confirm the difficulty of fully distinguishing between real effects and those arising from tax reporting when evasion is present.[3]

In this paper we investigate how people react to tax notches in a laboratory experiment. Differently from field studies, the experiment allows us to observe both real effort and income reporting. We implement a 2×2 experimental design which varies the type of tax system (proportional vs. notched) with the possibility of evasion. This allows us to address several questions on the effects of the tax notch system in the lab, including whether subjects' responses to notches are mainly driven by real adjustments or evasion adjustments, and whether the latter adjustments increase or decrease total evasion. For the first time to our knowledge, we can answer these questions based on direct experimental evidence.

We obtain several results. We find strong evidence of excessive bunching in the tax notch system, both with and without tax evasion possibility. The effort adjustments are mainly from above the threshold, while the evasion adjustments are mainly from below the threshold. Both adjustments contribute to a reduction in the underreporting rate, even if they do not affect the decisions to evade all income. We also confirm evidence

[2]Indeed, Kleven and Waseem (2013) for notches and Saez (2010) and Chetty et al. (2011) for kinks belong to a more general research program that uses bunching response to recover elasticity parameters representing sufficient statistics for welfare analyses (Chetty, 2009). See Slemrod (2013) and Lockwood (2020) for a discussion of the approach of sufficient statistics in tax systems with notches.

[3]In general, only indirect evidence of evasion can be revealed in empirical studies (Slemrod and Weber, 2012). For instance, several studies show the self-employed react more strongly to tax incentives than employees (e.g., Saez, 2010; Le Maire and Schjerning, 2013; Bastani and Selin, 2014; Miller et al., 2024; Kleven and Waseem, 2013; Adam et al., 2021; Hargaden, 2020). Likely, this arises from the greater opportunities for evasion among the self-employed, but employees' stronger frictions can also play a role. Tax evasion may be suggested by the tendency to report round numbers when calculating tax liability (Aghion et al., 2024). Over-reporting costs also remains a major evasion strategy in some cases (Best et al., 2015; Carrillo et al., 2023; Lobel et al., 2024), though sales adjustments without corresponding changes in input costs can also signal evasion (Di Marzio et al., 2024).

of optimisation frictions, which are stronger when tax evasion is possible than when it is not. A gender breakdown highlights both the robustness of the effects found and the large impact that heterogeneous preferences can have on the overall responses to tax notches.

Our experiment is in the vast literature on tax evasion in the laboratory (Alm and Malézieux, 2021). Starting from the seminal experiment on the tax evasion game by Friedland et al. (1978), lab experiments have contributed to the understanding of a wide range of issues in the analysis of tax evasion (comprehensive surveys on the tax compliance literature in Andreoni et al., 1998; Alm, 2012; Slemrod, 2019). Our experiment is mostly related to classical experiments on compliance response to tax policy variables, e.g., audit rates, penalty rates, and tax rates (Webley, 1987; Alm et al., 1992). We extend the analysis to the effects of tax notch policies, which are used in many more countries than previously recognized and are now increasingly the focus of public and professional attention.[4]

Few previous experiments have investigated the joint choices of effort and income tax reporting in the lab (Collins and Plumlee, 1991; Pántya et al., 2016). Closer to our setting, Doerrenberg and Duncan (2014) contrast effort choices with and without tax evasion possibilities in a proportional tax system. They find that the possibility of evasion mildly affect how people respond to changes in the tax rate, mainly when the tax rate decreases. We vary the existence of tax evasion possibilities in a notch tax system, which the empirical literature has shown to generate large behavioral responses, despite the difficulty of fully identifying whether due to real or evasion decisions.

Further, the experiment disentangles the effects of the tax notch threshold from other tax interventions that may affect responses to notches in the field. Indeed, as remarked by Kanbur and Keen (2014), real tax systems have multiple features that can markedly increase the range of notch responses and complicate optimal tax notch design.[5] Our experiment uniquely isolates taxpayers' responses due to pure tax notch effects.

The study also contributes to the literature on optimization frictions in the tax

[4]Notch policies are not only common in tax-benefit settings, but occur in several other areas of public interventions, such as business size rules (Kaplow, 2017; Gourio and Roys, 2014; Garicano et al., 2016), product certification policies (Sallee, 2014; Houde, 2018), international trade regulations (Fajgelbaum and Khandelwal, 2024), law enforcement policies (Traxler et al., 2018; Almunia and Lopez-Rodriguez, 2018; Hiraiwa et al., 2024).

[5]For instance, field tax notch policies are often accompanied by accounting simplifications, which may include exemptions from keeping books or issuing traceable receipts. While aimed at the same objective to reduce taxpayers' burden (in this case, compliance costs), these exemptions can nevertheless hinder effective audits and independently affect evasion decisions and bunching (Aghion et al., 2024; Di Marzio et al., 2024). (In this sense, accounting exemptions are similar to enforcement notches, conceptually distinct from tax notches, as they represent size-dependent enforcement policies; Almunia and Lopez-Rodriguez, 2018; Carrillo et al., 2017; Hungerman, 2023).

systems (Chetty et al., 2013; Kleven and Waseem, 2013; Adam et al., 2021). As far as we know, experimental evidence is limited to the contribution of Gibson et al. (2019), who tested the relation between tax notches and bunching. They confirmed that frictions in responding to notch incentives are not only due to adjustment costs or other field constraints but also depend on behavioral effects. Still, Gibson et al. (2019) did not introduce tax evasion possibilities. Our experiment contributes to exploring the contexts and possible multiple reasons for suboptimal responses to tax incentives when evasion is possible (Slemrod, 2019).

Finally, the experiment intersects with the literature on gender differences in economic behaviour, covering a wide range of domains (Croson and Gneezy, 2009). While our focus on gender differences is primarily conducted as a robustness check to examine heterogeneity in tax notch responses, our findings nevertheless provides novel evidence on gender responses to economic incentives, particularly created by notch thresholds.

The rest of the paper is organized as follows. Section 2 presents the theoretical framework for the experiment. Section 3 outlines the design, and Section 4 reports the results. Section 5 contains the concluding discussion.

2 Theoretical framework

We examine experimentally decisions on effort and income tax reporting. Pencavel (1979) presented the first model of these joint decisions, extending the original Allingham and Sandmo (1972) analysis of tax evasion with exogenous income. We introduce a tax notch and assume a quasi-linear preference model with evasion costs, a common approach in empirical studies (e.g., Saez, 2010).

A taxpayer makes an effort L to earn a gross income. Since the wage rate is 1, L directly corresponds to the gross income obtained. There is a tax schedule $T(L)$, with a notch at the threshold S. In particular, a preferential tax rate t_p is applied for income L not exceeding S, while a standard proportional tax rate t, with $t > t_p$, applies when L is greater than S. Thus, $T(L) = t_p L$ for $L \leq S$, and $T(L) = tL$ for $L > S$.

The tax authority doesn't know the taxpayer's gross income but audits what is reported with probability p. The taxpayer's reported income, X, can be any value within the range $[0, L]$, with $T(X)$ representing the tax paid on the reported amount. If an audit occurs, the true L is discovered, and the taxpayer is required to pay a fine on the unpaid tax. The fine is given by $F(L, X) = f[T(L) - T(X)]$, where $f > 1$ is the penalty rate. Thus, the net income y available for the taxpayer's consumption is uncertain and is given by: $y^c = L - T(X) - F(L, X)$ if, with probability p, the taxpayer

is caught evading; and $y^{nc} = L - T(X)$ if, with probability $(1-p)$, the taxpayer is not caught.

The taxpayer must decide how much effort L to produce and how much income X to report. We consider a model with quasi-linear preference and costs of evasion:

$$EU(L, y) = (1-p)y^{nc} + py^{c} - v(L) - K(L-X) \tag{1}$$

where the function $v(L)$ gives the disutility to work, with $v(0) = 0$, $v' > 0$ and $v'' > 0$; and where the function $K(L-X)$ captures the costs of underreporting. The model is of the same class used in recent empirical and field experimental studies (Saez, 2010; Kleven et al., 2011). It is based on some simplifications and some extensions compared to the traditional approach (Allingham and Sandmo, 1972; Pencavel, 1979). The quasi-linear specification restricts the effect of taxes on labour supply to the substitution effect. The cost function $K(X-L)$ encompasses various costs associated with evasion. In field studies, it may include administrative and psychological costs due to moral constraints or even of the risk of evasion neglected by the quasi-linear preference.[6] Our experiment does not include administrative costs, but the various psychological costs can influence the participants in the laboratory.

We allow for both variable and fixed psychological costs: $K(L-X) = k(L-X) + q1(X < L)$. The indicator function shows that the fixed cost $q > 0$ is paid whenever $X < L$. In the variable component, it is assumed $k(0) = 0$, with $k' > 0$ and $k'' > 0$.

The model's predictions are obtained by maximising equation (1) with respect to L and X. It is useful to start from the predictions for a pure proportional system without the notch, where the standard rate t applies to any $L > 0$. In this case, the first-order conditions are given by:

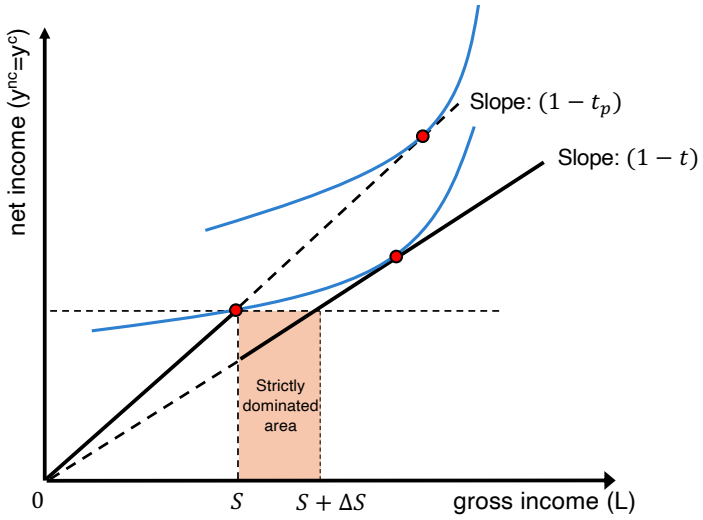
$$v'(L) = (1-t) \tag{2}$$

$$k'(L-X) = t(1-pf) \tag{3}$$

The conditions also indicate possible corner solutions. First, choosing not to work is never optimal (provided $v'(0) = k'(0) = 0$), while L tends to maximum potential as preferences approach Leontief. When the expected benefit of evading one euro of tax is

[6] However, this way of having convexity in the evasion decision differs from the expected utility model with risk aversion (Allingham and Sandmo, 1972; Yitzhaki, 1974). One difference is that the cost function in model (1) depends on the hidden income and includes the costs incurred in both audit and non-audit events, while the tax penalty (which depends on the unpaid tax) enters linearly. This implies that the quasi-linear specification (1) eliminates the substitution effect of a tax rate change on the reporting decision. This leaves only the income effect, which determines an intuitive negative relationship between the level of the tax rate and the decision to evade (see below).

Figure 1: Decision to bunch the effort under certainty.



strictly positive (i.e., $1-pf>0$), equation (3) (with $k(0)=k'(0)=0$) indicates that for a taxpayer to report their income fully, there must be a sufficiently high fixed cost of evasion ($q>0$) to make the taxpayer's utility at $X<L$ lower than at full compliance where $X=L$. Full reporting also occurs when evasion is not possible, for instance in the real world due to third party-reporting or in one of our experimental treatment by-design. When evasion is possible, full underreporting occurs when the marginal cost to evade in equation (3) is less than its benefit, that is $k'(L-X)<t(1-pf)$ any $L>X$. Also remark that the internal solution (2) for effort does not depend on the internal condition (3) for reported income. When it is optimal to report $X=0$, optimal effort can increase marginally from the internal solution in condition (2).

The tax notch affects the solution of the model. We begin by considering the case of full-reporting $X=L$, either because evasion is not possible or because of a high fixed cost of underreporting.[7]

2.1 Full reporting

With full reporting, the taxpayer's problem is only to choose effort, governed by first-order condition (2) in the proportional tax system. This is similar to the case considered in the empirical literature initiated by Kleven and Waseem (2013).[8] The effect of the

[7] In the following subsection, we include taxpayers whose underreporting cost q is so high that they fully report under the proportional tax system and continue to fully report under the notch system. The possibility of switching from full reporting to underreporting is considered in subsection 2.2 below.

[8] Nevertheless, a difference is that Kleven and Waseem (2013) analyze the effect of the tax notch by comparing it to a pure proportional tax system with the low tax rate (t_p in our framework), implying

tax notch can be illustrated graphically in Figure 1.

The notch corresponds to the discrete jump at the threshold S reducing the tax rate from t to t_p. This creates an area of strictly dominated effort decisions, corresponding to $(S, S + \Delta S]$ in the graph, where net income is greater and effort is lower than at the notch. The strictly dominated area does not depend on the decision makers' preferences nor on their productivity but only on the tax parameters. Its width ΔS is obtained by equating the net income at the threshold and at the upper bound of the strictly dominated area, namely $(S + \Delta S)(1 - t) = S(1 - t_p)$.

The dominated decisions may extend beyond the strictly dominated area, the more the taxpayer responds to the tax rate. The lower indifference curve in Figure 1 shows a marginal taxpayer, indifferent between staying at the notch or moving beyond the dominated region. In case of Leontief preference, the dominated region collapses to the strictly dominated one. An individual whose optimal decision at tax rate t is above the dominated region is unaffected by the notch. Conversely, a taxpayer whose optimal decision at rate t is within the dominated region should stop the effort at the threshold. In a population of taxpayers, this should cause many effort decisions to bunch at the threshold, creating a hole in the effort distribution, with the strictly dominated area being completely empty. Effort decisions in the dominated area can be due to optimization frictions. Kleven and Waseem (2013) found very large frictions, with about 90% of wage earners and between 50% to 80% of self-employed individuals in the dominated region unresponsive to notches. In the field, frictions can be due to adjustment costs (Adam et al., 2021; Tazhitdinova, 2020). They may also be due to inattention and cognitive bias. In the laboratory, Gibson et al. (2019) confirmed significant frictions and also found that behavioral factors can affect the propensity to bunch.[9]

Taxpayers with effort in the standard regime t below the threshold may respond to the preferential rate t_p by marginally increasing their effort (depending on the elasticity of effort with respect to the tax rate) and, in the limit, even contributing marginally to bunching from below at the threshold.

---

that adjustments to the notch are only from above the threshold. This is a useful identifying restriction for estimating structural elasticities. In our experiment, we are not interested in such a restriction. Instead, particularly when considering the effect of tax evasion, we are interested in investigating the impact of the introduction of a preferential tax notch regime on a standard regime, with adjustments that can occur both from above and from below the threshold.

[9] Mainly, Gibson et al. (2019) found that self-control and loss aversion increase bunching while overconfidence reduces it.

2.2 Underreporting

With underreporting, both L and X can respond to the tax notch. The optimal choice needs to compare the solutions at the first-order conditions with decisions where one or both choice variables are at the threshold S.

For taxpayers who can earn at most the threshold, the notch is equivalent to a reduction in the tax rate compared to the proportional system, from t to t_p. Given the first-order conditions (2) and (3), this decrease can produce an increase in both effort (in cases of positive elasticity) and reported income (since $k'' > 0$).

For taxpayers able to earn more than the threshold, adjustments depend on evasion in the proportional system. Taxpayers choosing $L \geq X \geq S$ under the proportional tax (i.e., fully compliant or evading but reporting at least S) have to check whether such solutions remain optimal with the notch or if one of the following is better: i) choosing both L and X at S; or) reducing X to S, and choosing L accordingly.[10] The first case is consistent with a full real adjustment and full reporting, with a reduction of the underreporting rate (at least, for the taxpayers not already fully reporting); whereas the second substitutes a real adjustment with a reported income adjustment, with a likely increase in evasion with respect to the proportional system.[11]

Taxpayers choosing $L > S$ but $X < S$ under the proportional system also have two alternatives. They can choose option i) (the same as above), where both L and X are at S; or option iii) where they can adjust L and increase X to at most S (bunching reported income from below).[12] Both of these alternatives lead to a reduction in underreporting, both absolute and relative, compared to the proportional tax without the notch.

It is worth noting that the above discussion also indicates that, in the case of underreporting, there isn't a dominated region in the notch system that applies to all evaders, regardless of their preferences.[13] Therefore, only taxpayers who decide to comply fully

[10] To be precise, this case requires solving the first-order condition from model (1) when $X = S$, given by $v'(L) = (1 - fpt) - k'(L - S)$. This leads to a choice of effort L that may be higher or lower than in condition (2), with the difference vanishing as preferences approach Leontief.

[11] Conversely, for a taxpayer with $L > X \geq S$ in the proportional system, choosing $L = S$ and $X < S$ in the notch system is unlikely to be optimal. This is because, if a taxpayer with L greater than S in the proportional system chooses to reduce it at S under the notch, then it would likely be more beneficial to comply fully and also eliminate the cost of underreporting.

[12] In this case, the first-order conditions for L and X are, respectively: (2') $v'(L) = (1 - fpt)$ and (3') $k'(L - X) = t_p(1 - fp)$, with a choice of effort L greater than in condition (2) (when effort elasticity is positive), and a choice of evasion $(L - X)$ lower than in condition (3).

[13] In particular, this is because the expected utility of any tax evasion gamble $(L > X)$ in model (1) always depends on the underreporting costs $K(L - X)$. Consequently, even for a tax evader whose first-order conditions (2) and (3) imply $L > X$, with $X \in (S, S + \Delta S]$ (strictly dominated region under certainty), it is always necessary to check whether reducing reported income at $X = S$ increases the underreporting cost more or less than the expected benefit.

should always empty the (strictly) dominated region.

2.3 Predictions

The above discussion is summarised in the following predictions for the responses to the tax nocth:

> **Effort.** The tax notch can generate bunching of effort at the threshold, both with and without tax evasion. If bunching occurs with full reporting, the strictly dominated region must be completely empty, with some further emptying possible above and below it, depending on how sensitive effort is to the tax rate in function v. If bunching occurs with underreporting, no single dominated choice region applies to all taxpayers who can bunch effort in the notch system from different effort levels in the proportional system.
>
> **Reported income.** With underreporting, reported income can be adjusted in addition to effort at the threshold. Reported income can be adjusted both from above and from below the threshold compared to the proportional system. While in the first case there is a likely increase in evasion, in the second case there is an increase in reported income with a reduction in the underreporting rate.

Finally, we note that the above predictions are similar to those obtained by Kanbur and Keen (2014) to propose a partition of taxpayers' behaviours in a notch tax system. One difference is that their model has no tax below the threshold (equivalent to setting $t_p = 0$ here).[14] Due to this feature, there is no reason in that model to report different false levels of income below the threshold, as all reports pay no tax and all individuals below the threshold are invisible - or 'ghosts' - to the tax authorities. In our model (and experiment), the positive t_p below S, even if very low, can still make a difference by inducing some taxpayers to report fully, some to report less than L but more than zero, and only individuals reporting and paying nothing can be considered genuine 'ghosts'.

3 Experimental design

We use the theoretical framework from the previous section to investigate the effect of a tax notch in a laboratory experiment. It consists of four treatments and employs a between-subjects design based on standard features from the experimental literature

[14] For this feature, they remark that while for clarity of exposition they refer to the tax base in their analysis as 'income', their model is particularly suited to value added taxes that almost everywhere have a notch created by 0 tax below some entry threshold level.

on tax evasion (Alm, 2012). All four treatments include the same real effort task, in which participants earn gross earnings over a series of periods. In each period, the gross earnings are subject to taxation. The four treatments differ in terms of the tax schedule and whether participants can underreport their earnings. Treatments 1 and 2 are based on a proportional tax system, with only treatment 2 participants that can underreport; while treatments 3 and 4 are based on a tax notch system, with only treatment 4 participants that can underreport.

3.1 Effort task and income reporting decisions

In all treatments, participants must complete individual effort tasks over 12 periods, the first two of which are trial rounds. The effort task consists of a computerised slider (Gill and Prowse, 2012, 2019). In each period, participants have 120 seconds to place as many as possible of 48 sliders on a screen at the center of a bar. For each correctly placed slider, a participant receives a constant payment of €1. During the slider task, subjects are kept informed of the number of sliders completed, the remaining time, their current gross income, and the tax regime which depends on the treatment. Following the slider task, in treatments where underreporting is possible, participants have 90 seconds to decide how much income to report. At the end of the period, before moving to the next, all subjects receive a 30 seconds feedback screen, showing the round's results, with the gross and net income earned, and whether they were audited if evasion was possible.

3.2 Treatments and tax schedule

The four treatments are shown in Table 1 with the tax schedules. Treatments 1 and 2 use a proportional tax rate of 40%. In treatment 1, subjects cannot underreport. In treatment 2, they can underreport, with the computer controlling their reported income with a probability of 10%. In the case of an audit, the total gross income is determined, and the participant is subject to a fine equal to twice the amount of unpaid tax. Treatments 3 and 4 have a notch scheme, with a reduced tax rate of 10% for income below a threshold, while the entire income is subject to the standard rate of 40% when it is above the threshold. In treatment 3, participants cannot underreport, while in treatment 4 they can, with the same audit probability and penalty as in treatment 2.

In both treatments 3 and 4, the threshold was set at €10, corresponding to 10 sliders, creating a discontinuity in the average tax rate and a strictly 'dominated region' between 11 and 15 sliders. Within this range, greater effort results in higher gross income but lower net income compared to that at the threshold. We chose this cut-off point based on

Table 1: The four between-subject treatments.

Treatment	Tax schedule	Tax rate	Enforcement	Participants (% females)	Periods
1. Proport_NO_Evasion	Proportional	40%	Evasion not possible	123 (52.0%)	10+2 trials
2. Proport_Evasion	Proportional	40%	Audit prob.: 10% Fine: $2 \times$ unpaid tax	129 (49.6%)	10+2 trials
3. Notch_NO_Evasion	Notch system	10% for income ≤ 10 40% for income > 10	Evasion not possible	125 (52.0%)	10+2 trials
4. Notch_Evasion	Notch system	10% for income ≤ 10 40% for income > 10	Audit prob.: 10% Fine: $2 \times$ unpaid tax	133 (54.9%)	10+2 trials

a pilot session of treatment 1, where the majority of participants were able to complete 10 sliders. In this way, most participants are affected by the presence of the threshold.

3.3 Payment and procedure

In all treatments, participants are paid according to the net income obtained in one of the periods between 3 and 12. The period is drawn at the end of the experiment by a participant from an envelope containing 10 tickets numbered 3 to 12.

Subjects are informed about the procedure at the start of the experimental sessions.[15] An experimenter reads aloud the instructions distributed to participants. They are told that they are about to participate in an individual choice experiment and that, in addition to a €3 show-up fee, the money they can earn depends on the performance in a computerized real effort task, which is explained in detail to subjects. Then, they are explained about the tax system of the various treatments, with tables distributed with the instructions listing the gross earnings, taxes due, and net incomes depending on the number sliders completed. In treatments 3 and 4 with the notch tax scheme, the dominated area is highlighted. In treatments 2 and 4, where underreporting is possible, subjects are informed about the auditing procedure and the fine in case of evasion.

At the end of the instructions, participants can ask questions, and before the experiment begins, they must answer few control questions to make sure they understand the experimental procedures. They cannot proceed with the experiment until they have answered all the control questions correctly. After all periods have been played, a short questionnaire asks participants about some demographic variables and asks them to self-rate their risk propensity on a ten-point scale.

[15] Instructions for the experiments and other materials, including the experimental screens and a picture of the experimental laboratory, are reported in Appendix A of the supplementary online material, along with the statistics of sample participants.

A total of 510 participants took part in the experiment, with 123 to 133 participants per treatment, fairly balanced between genders (Table 1). The participants were graduate and undergraduate students from the Ca' Foscari University of Venice, mainly from the fields of economics, management, languages and humanities. They were recruited using ORSEE.3 (Greiner, 2015). The experiment was pre-registered at AsPredicted.[16] Preregistration was for 90 subjects per treatment, with the experimental sessions taking place in December 2023. Additional sessions with new participants were run in February 2025 to increase the sample size for a detailed analysis of the gender breakdown. The experimental sessions were conducted at the VERA Laboratory for Experiments in Social Sciences (VERALabEx) of the university.

Experimental sessions lasted around 90 minutes (including time for reading the instructions aloud, answering questions, and paying), with participants earning an average of €13.50, including a €3 show-up fee. The experiment was programmed using zTree (Fischbacher, 2007).

4 Results

In this section we present the results of the experiment. We begin with an overview of the choice variables, looking at the main characteristics of the average behaviour across treatments. As there are differences in average behaviour between men and women, we present the evidence for the whole sample and separately by gender. We then turn to the key questions of the effect of the tax notch on bunching and emptying in the distribution of effort and reported income.

4.1 Overview of choice variables

Figure 2 shows the period averages of the choice variables and underreporting rates across treatments. The figure caption reports the averages for periods 3-12, which count for payment.[17] Starting with effort (measured by sliders completed and corresponding to gross income), the mean in the whole experiment is 16.42. The differences between treatments are small. The plot with the period averages in the four treatments also shows close trends and indicate a learning process in the slider task that becomes approximately linear from period 3 (the first valid for payment).

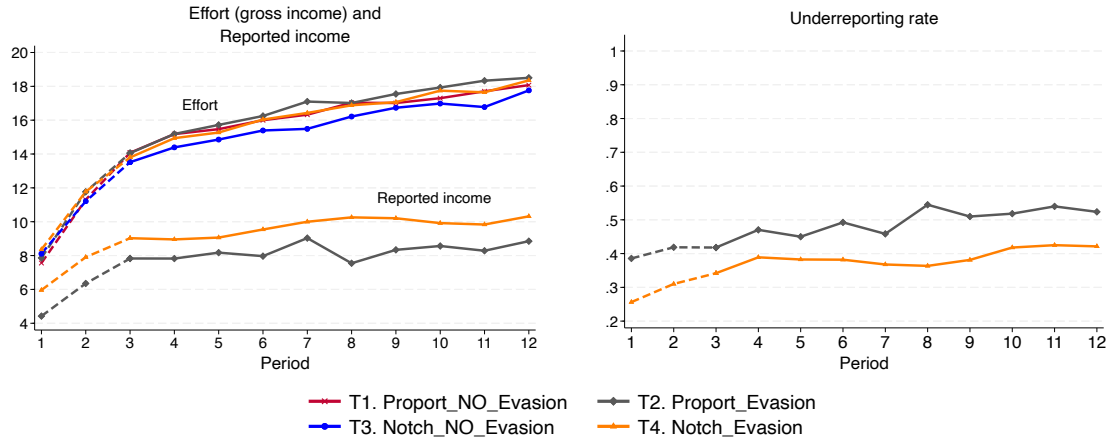
The differences in reporting behavior are more apparent. Both the period averages

[16] Aspredicted 150096 -'Effort or tax evasion with income tax notches: an experimental analysis', https://aspredicted.org/nyv8-mm6z.pdf.

[17] All statistics reported here and below refer to the ten payment periods.

and the overall averages in Figure 2 show less underreporting in treatment 4, with the tax notch, than in treatment 2, under the proportional tax (overall, about 10% less).

Figure 2: Average behavior across periods and treatments.



Overall averages and standard deviations (in parentheses) over payment periods 3-12 are: *Effort*: T1 ($n=123$): 16.42 (4.962); T2 ($n=129$): 16.76 (5.433); T3 ($n=125$): 15.81 (5.662); T4 ($n=133$): 16.42 (5.060); All ($n=510$): 16.36 (5.925). *Reported income*: T2 ($n=129$): 8.24 (7.032); T4 ($n=133$): 9.71 (6.965); T2 and T4 ($n=262$): 8.99 (7.035). *Underreporting rate*: T2 ($n=129$): 0.49 (0.395); T4 ($n=133$): 0.39 (0.394); T2 and T4 ($n=262$): 0.44 (0.398).

The regressions in columns (1)-(3) of Table 2 support the visual impression. They are conducted on the choice variables with standard errors clustered at the subject level and including a linear trend. Statistically significant differences are confirmed for the underreporting rate being higher in treatment 2 than in treatment 4 (col. 3) and reported income being lower (col. 2), while there is no significant difference between treatments for effort (col. 1).

Overall sample averages can be influenced by subjects' personal characteristics. Among other factors, previous research has shown that women tend to complete fewer sliders than men (Gill and Prowse, 2019). They also tend to be more compliant than men (Kastlunger et al., 2010; Alm and Malézieux, 2021). To isolate the effect of gender, the regressions in columns (4)-(6) of Table 2 include interaction terms between treatments and gender dummies, in addition to a linear time trend and other individual controls. Figure 3 reports period and overall averages by gender.

Both the averages and the regressions highlight the impact of gender. The evidence for effort confirms that men generally perform better than women in the slider task. However, there are differences between the proportional tax and the notch. Under proportional taxation, females complete on average 15.15 sliders in treatment 1, without the possibility of evasion, and 14.86 in treatment 2, with the possibility of evasion. For

Table 2: Regressions of choice variables on treatments and controls.

	(1) Effort	(2) Reported income	(3) Under-reporting rate	(4) Effort	(5) Reported income	(6) Under-reporting rate
T2. Proport_Evasion	0.346			0.764		
	(0.567)			(0.738)		
T3. Notch_NO_Evasion	-0.606			0.165		
	(0.583)			(0.868)		
T4. Notch_Evasion	-0.001	1.471*	-0.105**	-1.070	0.104	-0.072
	(0.530)	(0.729)	(0.041)	(0.873)	(1.135)	(0.057)
Female=1				-2.787***	-0.592	-0.091
				(0.783)	(1.082)	(0.058)
T2×Female=1				-1.087		
				(1.046)		
T3×Female=1				-1.532		
				(1.103)		
T4×Female=1				2.111*	2.636	-0.038
				(1.069)	(1.452)	(0.077)
Risk				0.085	-0.459**	0.032***
				(0.092)	(0.156)	(0.008)
Trend (linear)	0.440***	0.117**	0.009***	0.440***	0.117**	0.009***
	(0.017)	(0.039)	(0.002)	(0.017)	(0.039)	(0.002)
Constant	14.000***	7.600***	0.442***	14.341***	11.711***	0.316**
	(0.386)	(0.526)	(0.032)	(0.991)	(1.900)	(0.098)
Controls	No	No	No	Yes	Yes	Yes
Obs.	5100	2620	2620	5100	2620	2620
Clust.	510	262	262	510	262	262
F	174.27	6.59	11.50	48.08	2.91	5.60
Prob >F	0.0000	0.0016	0.0000	0.0000	0.0006	0.0000
R-squared	0.0610	0.0132	0.0219	0.1533	0.0655	0.1208

Notes. This table reports the coefficients from regressions of choice variables over periods 3-12 (payment periods). Standard errors are clustered at the subject level and reported in parentheses. For regressions in columns (1) and (4), the baseline is T1. Proport_NO_Evasion; for regressions in columns (2)-(3) and (5)-(6), the baseline is T2. Proport_Evasion. Controls in regressions (4)-(6) include age, student level, study field, and experience in previous experiments. (Estimates with all control coefficients are provided in the online supplementary material, Table B.2 of Appendix B.). Statistical significance indicated as: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Results of t-tests for the difference between responses from the regressions (in addition to those in the table) are below (standard errors clustered in parentheses).

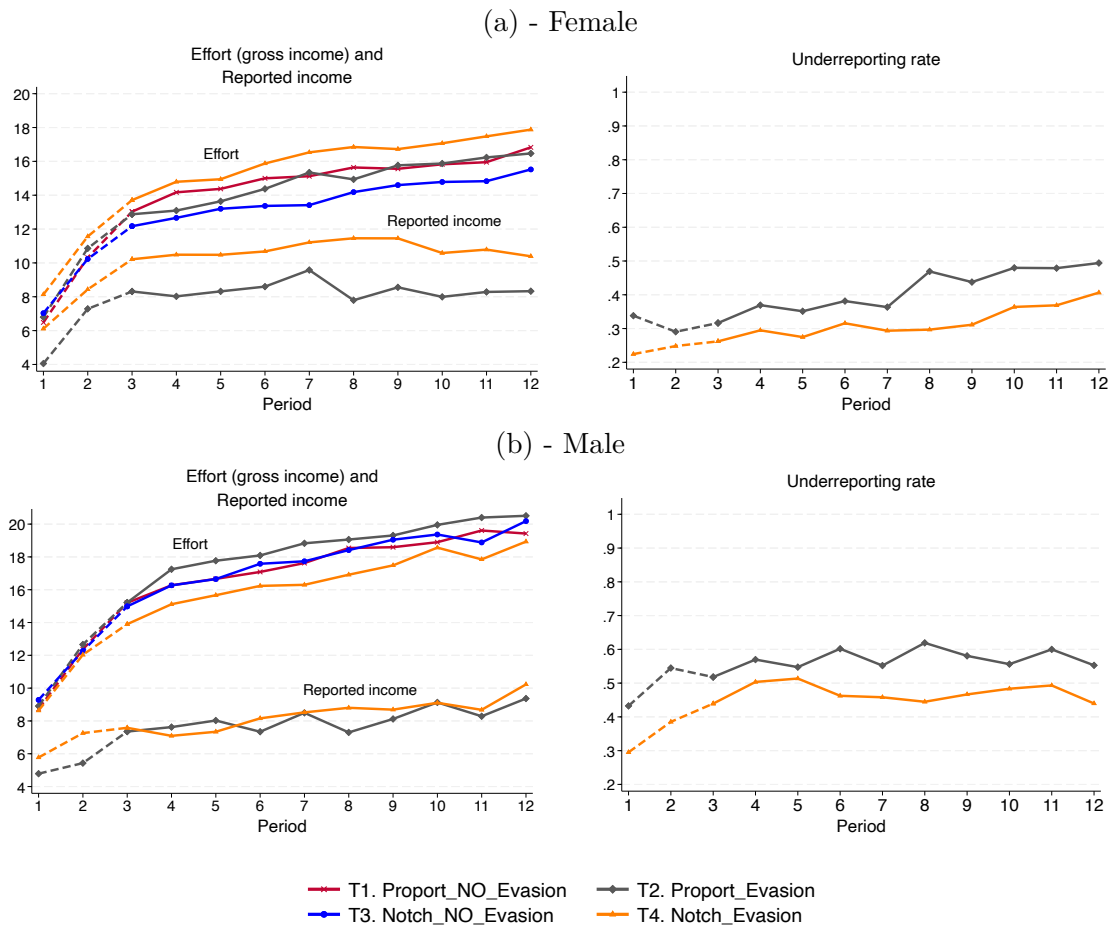
Regressions (1)-(3): Differences between treatments (whole sample). *Effort*: T3 vs. T2: −0.952 (0.605); T4 vs. T2: −0.348 (0.555); T4 vs. T3: 0.605 (0.570).

Regressions (4)-(6): Gender and treatment differences by gender. *Gender differences within treatments (Male - Female)*: *Effort*: T1: 2.787*** (0.783); T2: 3.874*** (0.763); T3: 4.319*** (0.824); T4: 0.676 (0.761). *Reported income*: T2: 0.592 (1.082); T4: −2.044 (1.067). *Underreporting rate*: T2: 0.091 (0.058); T4: 0.128* (0.057).

Female, treatment differences: *Effort*: T2 vs. T1: −0.323 (0.763); T3 vs. T1: −1.368* (0.690); T3 vs. T2: −1.044 (0.789); T4 vs. T1: 1.040 (0.644); T4 vs. T2: 1.363 (0.755); T4 vs. T3: 2.407*** (0.679). *Reported income*: T4 vs. T2: 2.741** (0.927). *Underreporting rate*: T4 vs. T2: −0.110* (0.051).

Male, treatment differences: *Effort*: T2 vs. T1: 0.764 (0.738); T3 vs. T1: 0.165 (0.868); T3 vs. T2: −0.599 (0.788); T4 vs. T1: −1.070 (0.873); T4 vs. T2: −1.834* (0.773); T4 vs. T3: −1.235 (0.877). *Reported income*: T4 vs. T2: 0.104 (1.135). *Underreporting rate*: T4 vs. T2: −0.072 (0.057).

Figure 3: Average behavior across periods and treatments, by gender.



Overall averages and standard deviations (in parentheses) over payment periods 3-12 are as follows:
(a) Females. *Effort*: T1 ($n=64$): 15.15 (4.462); T2 ($n=64$): 14.86 (5.503); T3 ($n=65$): 13.87 (4.980); T4 ($n=73$): 16.19 (5.060); All ($n=266$): 15.05 (4.981). *Reported income*: T2 ($n=64$): 8.38 (6.348); T4 ($n=73$): 10.78 (6.421); T2 & 4 ($n=137$): 9.66 (6.496). *Underreporting rate*: T2 ($n=64$): 0.41 (0.387); T4 ($n=73$): 0.32 (0.359); T2 & 4 ($n=137$): 0.36 (0.376).
(b) Males. *Effort*: T1 ($n=59$): 17.79 (5.114); T2 ($n=65$): 18.64 (4.662); T3 ($n=60$): 17.91 (5.611); T4 ($n=60$): 16.70 (5.484); All ($n=244$): 17.78 (5.264). *Reported income*: T2 ($n=65$): 8.11 (7.648); T4 ($n=60$): 8.42 (7.376); T2 and T4 ($n=125$): 8.26 (7.517). *Underreporting rate*: T2 ($n=65$): 0.57 (0.388); T4 ($n=60$): 0.47 (0.419); T2 and T4 ($n=125$): 0.52 (0.406).

males, the corresponding averages are 17.79 and 18.64. Regression analysis (column 4 of Table 2) confirms that while for both genders the differences between the two treatments are not statistically significant, the gender differences are highly statistically significant in both treatments. The gender difference increases further in treatment 3, with the notch and no possibility of underreporting, where the women's average decreases to 13.87, which is their lowest average effort, while for men it remains at about 17.91.

Interestingly, the gender difference becomes very small and not statistically significant in treatment 4, with the notch and the possibility of underreporting, where women's average is 16.19 and men's 16.70. Thus, treatment 4 obtains the highest effort from women and the least effort from men.

Also interesting are the gender differences in reporting behaviors. Women match the increase in effort between treatment 2 and 4 with an even higher increase in reported income, which leads to a statistically significant reduction in the underreporting rate of about 10% (from an average of 41% to 31%). Men's reported income in treatments 2 and 4 is about the same, but due to the reduced effort, the underreporting rate decreases between treatment 2 and 4 for males as well (from an average of 57% to 47%), although for males alone the reduction does not reach statistical significance at the conventional level in the regression in Table 2.

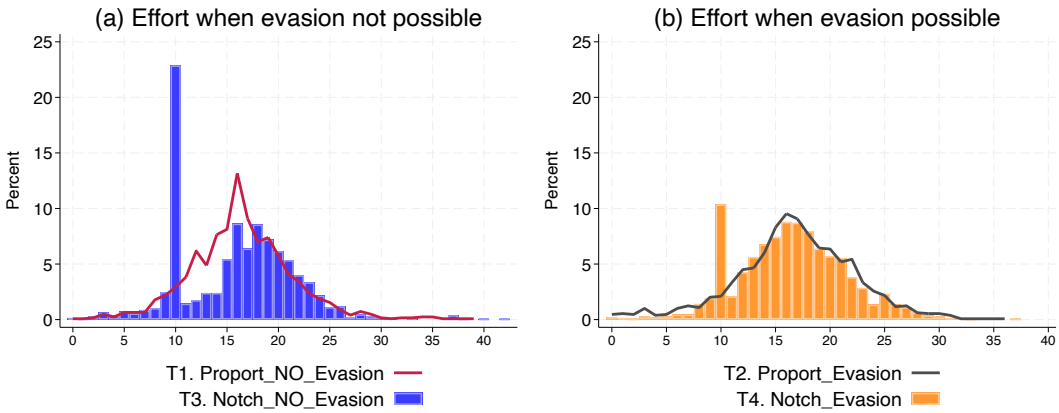
The above gender differences highlight important elements of heterogeneity and compositional effects arising from responses to tax notches. However, average effects offer a coarse explanation of these responses. As emphasized theoretically, responses to tax notches are primarily expected as bunching in the distributions of the choice variables. We now turn to the analysis of these distributions. We first look at the total subject pool and provide various results. We then conduct robustness checks on the main effects found, including a breakdown by gender. We begin the analysis of the effort adjustments.

4.2 Bunching of effort

Figure 4 compares, between the proportional and notch tax treatments, the distributions of effort when evasion is not possible (panel a) and when is possible (panel b). In both notch treatments, bunching at the threshold of 10 sliders stands out. It is very high when evasion is not possible (treatment 3), at about 24% of the total treatment choices. It is lower when evasion is possible (treatment 4), at about 10%. Also, in treatment 3, bunching creates a clear hole above the threshold. Conversely, the origin of bunching is less clear in treatment 4 and less concentrated in a specific area.

To assess the significance of the visual evidence, Table 3 reports the predicted probabilities and marginal effects estimated by a multinomial logit model for the effect of the four treatments on effort, which is classified into five relevant categories: (1) below the 10 threshold; (2) at the threshold; (3) in the strictly dominated area between 10 and 15 sliders; (4) in a slightly higher area between 15 and 18 (with its upper limit visually derived from panel (a) of Figure 4); and (5) above 18 sliders. The model is estimated with clustered standard errors at the subject level, including linear trends, personal controls and the treatment dummies interacted with gender. The estimates of the multinomial

Figure 4: Distributions of effort by treatments.



coefficients are reported in the online supplementary material. The gender breakdown is discussed in Section 4.6.

The predicted probabilities confirm an effort concentration at the 10 threshold in treatments 3 and 4 with the notch, while the proportions of choices at 10 are small in treatments 1 and 2 without the notch. The marginal effects show small differences across all effort categories between the proportional tax treatments 1 and 2, with or without tax evasion. Large and significant 'excess' bunching (col. 2) is found in treatment 3 compared to treatment 1 and in treatment 4 compared to both treatments 1 and 2, the latter with evasion under the proportional tax. Bunching is significantly stronger in treatment 3 than in treatment 4: about 13% fewer choices, corresponding to about 65% of more bunching observed in treatment 3. Overall, the evidence provides a neat result that is consistent with the theoretical model. It also confirms that the possibility to evade significantly reduces the impact on real effort adjustments.

R1. *Bunching of effort at the threshold occurs both with and without the possibility of evasion. It is stronger, more than twice as strong (24% vs. 10%), when tax evasion is not possible than when it is.*

4.3 Holes and optimisation frictions in the effort distributions

Next we turn to the question of which areas are emptied to create bunching in the notch treatments. In treatment 3, the estimates show a statistically significant reduction of the effort in the strictly dominated region (Table 3, col. 3). Nevertheless, the region does not empty completely and about 13% ($p < 0.001$) of choices remain in the region.

Table 3: Predicted probabilities and marginal effects of effort choices across treatments (Multinomial Logit).

	(1) Eff.< 10	(2) Eff. $= 10$ Bunching	(3) Eff. $(10-15]$ Str. domin.	(4) Eff. $(15-18]$	(5) Eff. > 18
Predicted probabilities					
T1. Proport_NO_Evasion	0.07	0.03	0.30	0.30	0.30
	(0.015)	(0.006)	(0.023)	(0.020)	(0.029)
T2. Proport_Evasion	0.09	0.02	0.27	0.26	0.35
	(0.019)	(0.005)	(0.021)	(0.019)	(0.029)
T3. Notch_NO_Evasion	0.07	0.24	0.13	0.23	0.33
	(0.015)	(0.026)	(0.014)	(0.018)	(0.027)
T4. Notch_Evasion	0.06	0.10	0.26	0.25	0.33
	(0.011)	(0.017)	(0.021)	(0.019)	(0.030)
Marginal effects					
T2 vs. T1	0.020	-0.008	-0.030	-0.035	0.053
	(0.024)	(0.008)	(0.031)	(0.028)	(0.040)
T3 vs. T1	0.002	0.207***	-0.172***	-0.064*	0.028
	(0.022)	(0.027)	(0.027)	(0.027)	(0.040)
T4 vs. T1	-0.015	0.072***	-0.043	-0.045	0.030
	(0.019)	(0.019)	(0.033)	(0.028)	(0.042)
T4 vs. T2	-0.034	0.080***	-0.013	-0.010	-0.023
	(0.022)	(0.018)	(0.030)	(0.027)	(0.042)
T4 vs. T3	-0.016	-0.134***	0.129***	0.020	0.002
	(0.019)	(0.032)	(0.025)	(0.026)	(0.041)

Notes. This table reports the predicted probabilities and marginal effects of effort choices in the four treatments (510 clusters, 5110 observations). The Multinomial Logit model used to derive the results is estimated with the treatment dummies interacted with gender, a linear time trend, and personal controls (estimated coefficients are provided in the online supplementary material, Table B.3 of Appendix B). Standard errors clustered at the subject level are in parentheses. Stars for statistical significance are only reported for marginal effects: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

This confirms the presence of optimization frictions in the lab (Gibson et al., 2019). In treatment 3, there is also some effort reduction above the strictly dominated region (col. 4).[18] This may be due either to an elasticity effect or to the anxiety of some subjects about not completing enough sliders to move out of the strictly dominated area. The fact that there is no change between treatments 1 and 3 below 10 sliders (col. 1) suggests that the elasticity effect may be small after all.

[18]The effect becomes not significant if the upper limit of category (4) is extended beyond 18 sliders.

In treatment 4, effort in the strictly dominated region accounts for about 26% of decisions (col. 3, Table 3). This is slightly less than in treatments 1 and 2 without the notch, but the difference is not statistically significant. The comparison with treatment 3 shows twice as many frequencies, with the difference being highly significant. Nevertheless, as noted in the theoretical section, the strictly dominated region should be emptied only by taxpayers who fully comply, while subjects who evade may still find it optimal to make an effort in the region.

Table 4: Predicted probabilities and marginal effects of bunching and strictly dominated effort choices in treatment 4 (Logit models).

	(1) Bunching Eff. $= 10$	(2) Str. domin. Eff. $(10 - 15]$
Predicted probabilities		
FC=1	0.26	0.27
	(0.039)	(0.034)
FC=0	0.04	0.26
	(0.010)	(0.023)
Marginal effects		
T4: FC=1 vs. FC=0	0.220***	0.009
	(0.040)	(0.039)

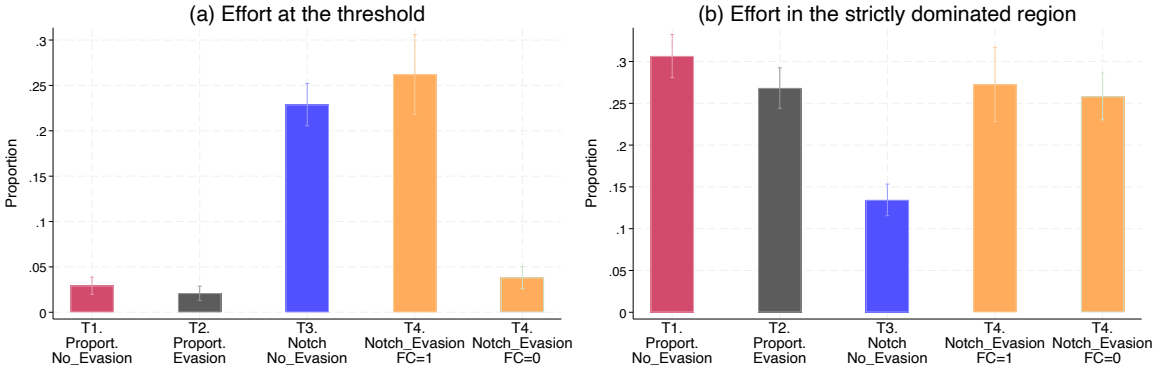
Notes. This table reports the predicted probabilities and marginal effects of bunching and strictly dominated effort choices in treatment 4 (133 clusters, 1330 observations). FC=1 indicates choice with full compliance, while FC=0 indicates choice with evasion. The Logit models used to derive the results are estimated with a gender dummy, a linear time trend, and personal controls (estimated coefficients are provided in the online supplementary material, Table B.4 of Appendix B). Standard errors clustered at the subject level are in parentheses. Stars for statistical significance are only reported for marginal effects: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

To clarify bunching and dominated effort choices in treatment 4, Table 4 reports predicted probabilities and marginal effects, derived from two logit models, that distinguish between full compliance (FC=1) and positive evasion (FC=0).[19] The same breakdown is shown visually in Figure 5, also showing a comparison with the other treatments.

[19]We also estimated the conditional probabilities for the choices of effort in treatment 4 using a bivariate probit approach with sample selection to control for the potential endogeneity of the choice between full compliance and evasion. We found no statistically significant impact in either the analysis of bunching or for effort in the strictly dominated area, with the Wald test of independent equations indicating no selection correlation. This may not be so surprising given the theoretical model in Section 2, where the decision at the extensive margin on whether to fully comply or not mainly depends on the level of the fixed psychological costs.

Effort margins at the threshold (col. 1, Table 4) support the prediction that bunching in treatment 4 is primarily associated with full compliance, given no advantage in treatment 4 in limiting effort and underreporting. Consistent with this, the small proportion of effort at 10 with positive evasion (0.04) is also similar to that in treatments 1 and 2 under the proportional tax (see Figure 5, panel (a)), while a large difference exists when comparing these treatments with treatment 4 under full compliance. The effects are symmetrical between treatment 4 and treatment 3, where subjects cannot evade and all have an incentive to bunch.

Figure 5: Bunching of effort and emptying of the strictly dominated area.



Results in the strictly dominated area (col. 2, Table 4) are less consistent with predictions. Whereas treatment 4 with full compliance should show no choices in this region, many choices are observed, with almost identical frequencies for full compliance (27%) and evasion (26%). These proportions are also similar to those in treatments 1 and 2 without the tax notch (Figure 5, panel (b)), while they double ($p<0.001$) those in treatment 3 with the notch but without evasion. Different cognitive effects could contribute to explain the evidence. One possibility is that in treatment 4 the dual decision of effort and tax reporting increases subjects' cognitive load and the chance of suboptimal choices compared to treatment 3, where there is only one decision. Another effect could be related to forms of dynamic inconsistency and myopia. With evasion possible, subjects in treatment 4 might try harder to move beyond the dominated area, viewing evasion as a possible insurance against the risk of remaining in the strictly dominated area. Frictions could arise if they fail to move beyond the area but still fully comply, perhaps due to underestimation of the moral constraint or fear of detection.

The evidence on holes and optimisation frictions is summarised.

R2. *There is emptying of the dominated region when evasion is not possible, even if*

optimisation frictions (13% of all choices) prevent it from emptying completely. When evasion is possible, emptying is less concentrated in any region above the threshold. Furthermore, optimisation frictions are stronger (27% of choices with full compliance).

4.4 Reported income adjustments

We move on to the analysis of the reported income adjustments. Figure 6 compares the distribution of reported income between treatments 2 and 4, both with the possibility of evasion, the former with the proportional tax and the second with the notch. The comparison shows a high concentration of reported income at the 10 threshold in treatment 4, at about 24%. When compared to the bunching of effort in the same treatment (about 10%, col. 2 of Table 3), this concentration clearly attests to an effect due to reported income adjustments rather than effort adjustments. Figure 6 also indicates an area of emptying in the distribution of reported income in treatment 4 below the threshold, whereas above the threshold there are no clear differences with treatment 2.

Figure 6: Distributions of reported income in treatments 2 and 4.

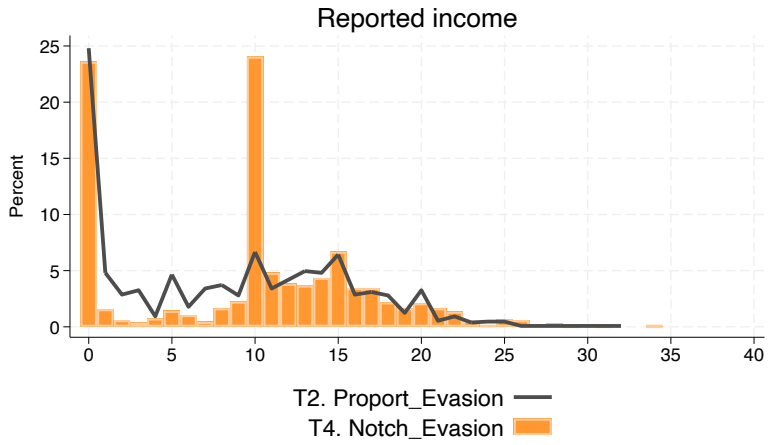


Table 5 disentangles in more detail the adjustments occurring between the two treatments. It reports the predicted probabilities and marginal effects estimated by a multinomial logit model for eight categories of joint effort and reported income decisions. In particular, the model distinguishes between the choices of effort: 1) below the threshold with full compliance; 2) below the threshold with evasion; 3) at the threshold with full compliance; 4) at the threshold with evasion; 5) above the threshold with full compliance. For effort above the threshold with evasion, the model further distinguishes choices based on reported income: 6) below the threshold; 7) at the threshold; 8) and above the

threshold. As in the previous models, the estimates use the respondents' choices over periods 3-12 as the cluster unit of analysis, with treatment dummies interacted with gender, time trend and personal controls.

The main effects occur for: effort at the 10 threshold with full compliance (col. 3); effort above 10 with reported income less than 10 (col. 6); and effort above 10 with reported income at 10 (col. 7). Summing the first and third effects, we can calculate about 16% of reported income bunching in treatment 4 in excess of treatment 2. Of this, approximately 43% (0.069/(0.069+0.09)) is due to effort adjustments and 57% to tax evasion. Facing this bunching of reported incomes at the threshold, there is a large reduction in effort choices above the threshold with reported income below the threshold (col. 6, -17%). The other marginal effects show virtually no change between the two treatments, in particular, no reduction in treatment 4 for decisions where both effort and reported income are above the threshold, either with full compliance (col. 5), or with evasion (col. 8). This confirms that no bunching of reported income occurs from above the threshold without also a real adjustment. Moreover, there is no increase of evasion when effort is below (col. 2) or at the threshold (col. 4). Overall, this indicates no reported income at the threshold with increased evasion. Instead, bunching of reported income occurs with reduced underreporting, driven by either more full compliance at the threshold (col. 3) or less evasion below the threshold with increased reported income at the threshold (col. 7).

R3. *When evasion is possible, approximately 43% of the excess bunching of reported income is due to effort adjustments and 57% to tax evasion adjustments. Effort adjustments occurs mainly with an increase of full compliance, and reported income adjustments with a decrease of evasion below the threshold. Both adjustments are consistent with a reduction in the underreporting rate in the experiment (cf. Table 2 above).*

4.5 Underreporting rates

Here, we better investigate the impact of the notch on underreporting. Figure 7 shows the distributions of underreporting rates in treatments 2 and 4, confirming a shift in treatment 4 towards more reporting, primarily full compliance. The effect appears largely due to a reduction in medium-large evasion (underreporting rates $> 50\%$), but with little effect on the choice of total evasion.

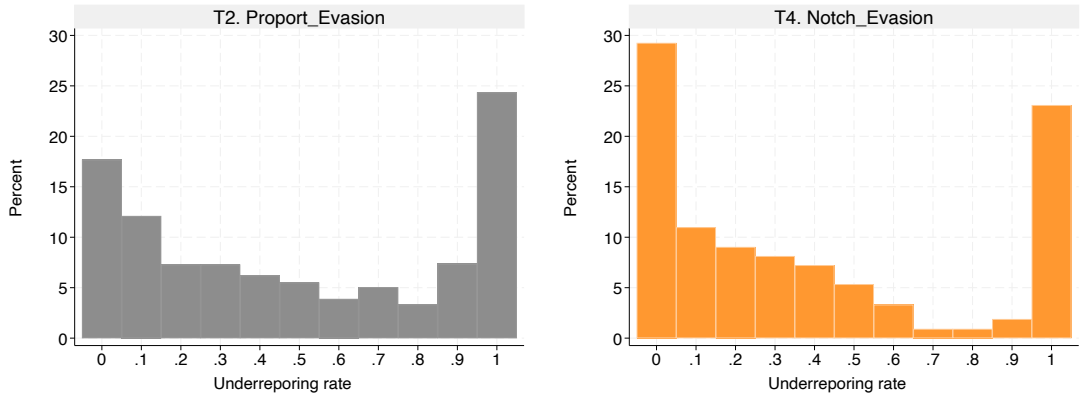
The visual impression is supported by Table 6, which reports the predicted probabilities and marginal effects between treatment 4 and treatment 2. These are estimated with a multinomial logit model for the underreporting rates in four categories: (1) full

Table 5: Predicted probabilities and marginal effects of joint reported income and effort choices in treatments 2 and 4 (Multinomial Logit).

	(1) Eff.<10 FC	(2) Eff.<10 Ev	(3) Eff.=10 FC	(4) Eff.=10 Ev	(5) Eff.>10 FC	(6) Eff.>10 Rep.<10	(7) Eff.>10 Rep.=10	(8) Eff.>Rep. Rep.>10
Predicted probabilities								
T2. Proport_Evasion	0.03	0.07	0.01	0.02	0.15	0.43	0.05	0.25
	(0.008)	(0.015)	(0.002)	(0.005)	(0.023)	(0.029)	(0.008)	(0.026)
T4. Notch_Evasion	0.03	0.03	0.07	0.03	0.18	0.26	0.14	0.26
	(0.007)	(0.006)	(0.015)	(0.007)	(0.022)	(0.030)	(0.021)	(0.029)
Marginal effects								
T4 vs. T2	-0.000	-0.040*	0.069***	0.013	0.023	-0.170***	0.090***	0.015
	(0.011)	(0.016)	(0.016)	(0.008)	(0.032)	(0.042)	(0.023)	(0.040)

Notes. This table reports the predicted probabilities and marginal effects of joint reported and effort choices in treatments 2 an 4 (262 clusters, 2620 observations). In columns (1), (3) and (5), FC indicates choice with full-compliance; in columns (2) and (4), Ev indicates choice with evasion; in columns (6)-(8) all choice are with evasion for different levels of reported income. The Multinomial Logit model used to derive the results is estimated with the treatment dummies interacted with gender, a linear time trend, and personal controls (estimated coefficients are provided in the online supplementary material, Table B.5 of Appendix B). Standard errors clustered at the subject level are in parentheses. Stars for statistical significance are only reported for marginal effects: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Figure 7: Distributions of the underreporting rates in treatments 2 and 4.



compliance, (2) positive evasion rates no greater than 50%, (3) evasion rates above 50% but less than 1, and (4) total evasion (underreporting rate = 1). There is a statistically significant decrease in large-scale evasion, largely offset by an increase in full compliance. Choices of total evasion remain virtually unchanged.

Table 6: Predicted probabilities and marginal effects of underreporting rates (URR) in treatments 2 and 4 (Multinomial Logit).

	(1) Full compliance	(2) Medium-small evasion (0 <URR≤ 0.5)	(3) Medium-large evasion (0.5 <URR< 1)	(4) Total evasion
Predicted probabilities				
T2. Proport_Evasion	0.19	0.36	0.21	0.24
	(0.024)	(0.029)	(0.026)	(0.029)
T4. Notch_Evasion	0.28	0.40	0.09	0.24
	(0.028)	(0.032)	(0.016)	(0.029)
Marginal effects				
T4 vs. T2	0.092*	0.033	-0.122***	-0.003
	(0.038)	(0.044)	(0.031)	(0.041)

Notes. This table reports the predicted probabilities and marginal effects of underreporting rates in treatments 2 and 4 (262 clusters, 2620 observations). The Multinomial Logit model used to derive the results is estimated with the treatment dummies interacted with gender, a linear time trend, and personal controls (estimated coefficients are provided in the online supplementary material, Table B.6 of Appendix B). Standard errors clustered at the subject level are in parentheses. Stars for statistical significance are only reported for marginal effects: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Two observations are worth making about these results. First, the evidence accords with the intuition underlying the quasi-linear preference model, with psychological costs of evasion, as outlined in Section 2. The model implies that these costs represent the

main force driving a taxpayer to comply with taxes, even partially, while the absence or very low level of these costs is the primary reason for total evasion. Since the tax notch doesn't affect these costs, it's not entirely unexpected that the choices of total evasion don't change with the notch. From a policy perspective, however, the evidence suggests that the notch system with a preferential tax rate below the threshold might affect large-scale tax evasion. However, its impact on the total black economy and tax 'ghosts' is likely to remain minimal.

R4. *The increase in tax compliance under the tax notch system is mainly due to a reduction in the medium-high evasion rates. The impact on total evasion decisions is minimal.*

4.6 Breakdown of adjustments by gender

This section returns to the gender differences highlighted in Section 4.1. Despite a balanced sample, it is important to assess the robustness of the results across genders and understand the origin of these differences.

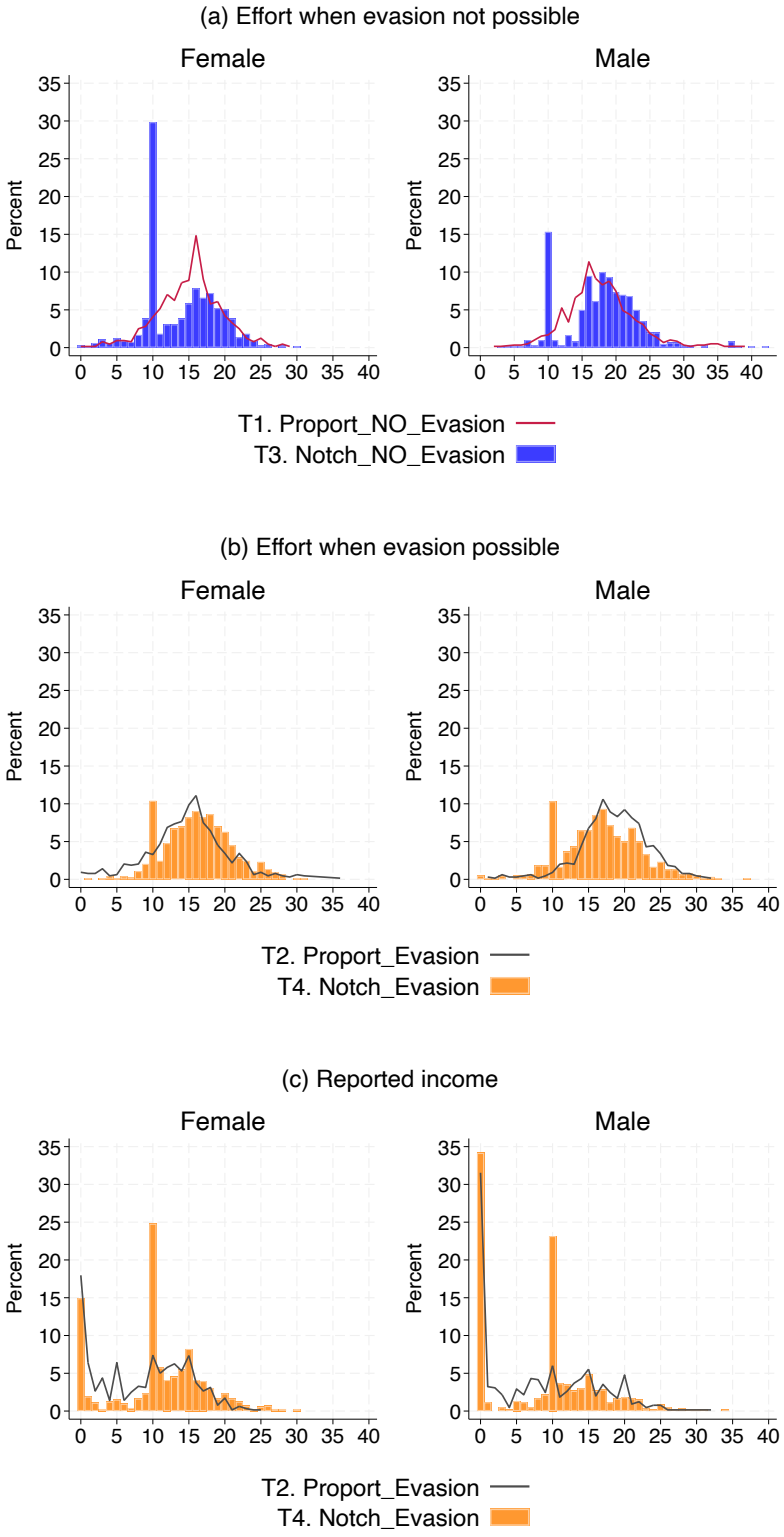
Figure 8 shows the distributions for males and females of the choice variables in the experimental treatments.[20] Treatments 3 and 4 show clear bunching at the threshold for both genders: of effort without and with evasion (panels a and b, respectively), and of reported income (panel c). There are also differences that are worth noting.

When evasion is not possible (panel a), bunching of effort in treatment 3 is significantly higher for females than for males (+14%, $p<0.01$). Since in the baseline treatment 1, with no tax notch and no evasion, females complete on average fewer sliders than males (-2.787 sliders, $p<0.001$; cf. Section 4.1) and have an average effort very close to 15 sliders, this can explain why in treatment 3 they tend to bunch more than males in order to avoid the strictly dominated area of the notch system (which has an upper limit exactly at 15 sliders).

When evasion is possible (panel b), bunching of effort in treatment 4 is quite close between males and females, with a frequency peak at 10% for both genders. However, when comparing with treatment 2, females' distribution in treatment 4 is slightly to the right, whereas males' is more to the left. A possible explanation for this difference is the relative strength of the two mechanisms behind bunching of effort found in the experiment when evasion is possible: the fear of ending up in the strictly dominated region; and the possibility of taking advantage of the low tax rate rather than bearing

[20] The detailed statistics on the gender breakdown are available in the online supplementary material, Appendix C.

Figure 8: Distributions of effort and reported income, by gender.



the costs of underreporting. For the reason just noted in treatment 3, females may feel more pressure from the first mechanism, namely the risk of falling into the dominated area. However, with evasion possible, they may try harder on the slider task, perhaps even considering underreporting if they do not succeed. This can explain the rightward shift of their effort distribution in treatment 4, which reports the highest average effort of females across all treatments (cf. Figure 3).

Males are generally better at the slider task; hence, they are on average less exposed to the risk of falling in the dominated area. For them, it is the second mechanism that can be relatively more important in the decisions to bunch effort at the tax notch, which leads to the leftward shift in their effort distribution in treatment 4, with a fall in the overall average (cf. Figure 3).

These varying intensities of the effort responses between men and women are relevant. They demonstrate the significant impact that heterogeneity among people can have on the effect of tax notches. A specific point highlighted is the possible role of people's self-confidence in the ability to cope with thresholds and dominated regions, and how tax evasion might impact this. It is certainly striking that treatment 4 generates for females the highest average effort, with the lowest underreporting rate.[21]

Nevertheless, despite differences in intensity, the evidence on the gender breakdown shows that all the main results of the previous sections are confirmed for both genders. These include the areas of emptying in the distributions of reported income, showing that evasion adjustments to the notch occur mainly from below the threshold (panel c, Figure 8). Moreover, it is shown in more detail in the online supplementary material that both males and females are exposed to optimizations frictions, which are stronger when evasion is possible; and that for both females and males, the tax notch contributes to a reduction in the medium-high evasion rates, but not of total evasion. It is on these general consistencies that we emphasize the overall evidence of the gender breakdown.

R5. *The responses of women and men to the tax notches are affected by noteworthy differences in intensity, mainly concerning effort with and without tax evasion. Never-*

[21] In the literature, there are two main streams that can be related to the results on gender differences. First, studies show that women often perform worse than men under pressure, including time constraints (as in the 120-second slider task). Yet, with relaxed constraints, their performance often increases, matching men's (Shurchkov, 2012; Galasso and Profeta, 2024; Buser et al., 2025). Thus, the possibility of evasion in treatment 4 might act as a relaxation of this time constraint, improving female performance. Second, a mechanism possibly operating jointly could be reference-dependent effects (discussed regarding tax notches by Kleven, 2016, and Gibson et al., 2019). Taxpayers may use the strictly dominated region's limits as reference points. The perceived distance from these points, and whether one perceives a loss (e.g., females falling short of moving beyond the dominated region) or a gain (e.g., males feeling confident of exceeding the region), can trigger different behavioral responses.

theless, the overall patterns of the results are consistent, including evidence on bunching, tax evasion responses, and optimization frictions.

4.7 Robustness over periods

A second robustness check concerns the stability of the evidence across rounds. There is considerable evidence that subjects learn during experiments. Their decisions improve in performance and become less affected by behavioural biases. It is relevant to examine whether and, if so, how learning affects responses to the tax notch in our experiment.

To this end, Figure 9 shows the distributions of the choice variables in the four treatments, dividing the 10 payment periods into periods 3-9 and the last three periods 10-12.[22] Bunching at the thresholds for both effort and reported income remain clearly visible throughout the experiment, with the other areas most affected by the tax notch. There is also a rightward shift in the effort distributions for periods 10-12 with thinning in the lower tail, including in treatments 1 and 2 without the notch, signaling learning in the slider task.

One question of specific interest is the impact of learning on optimisation frictions in treatment 3 and 4. Figure 10 shows the proportions of effort decisions in the strictly dominated area (10-15 sliders) across treatments for the two period intervals. The figure also divides treatment 4 into choices with full compliance (FC=1) and evasion (FC= 0), the latter not necessarily corresponding to frictions. In all treatments, the proportions of sliders decrease in the last three periods. For treatments 1 and 2, this is mainly due to learning in the slider task.[23] The reduction is most apparent in treatment 3, where the proportion of effort within 10-15 sliders reduced by more than half between the two period intervals, from 0.16 to 0.07 (-0.09, $p<0.001$), which is the lowest proportion observed across all treatments. In treatment 4, the reductions are from 0.31 to 0.19 (-0.12, $p<0.05$) for choices with full compliance, and from 0.30 to 0.17 (-0.13, $p<0.001$) with evasion. Both reductions are consistent. Nevertheless, the proportions of choice in the region remains high, including with full compliance, and in line to the proportions in treatments 1 and 2. Overall, this indicates a greater persistence of optimization frictions when subjects can underreport.

[22]The online supplementary material (Appendix D) reports the results of all prior statistical analyses conducted for the two distinct period intervals.

[23]In treatment 1, the reduction is from a proportion of 0.34 to 0.23 (-0.11, $p<0.001$), and in treatment 2 it is from 0.29 to 0.21 (-0.08, $p<0.01$).

Figure 9: Distributions of effort and reported income - Periods 3-9 and 10-12.

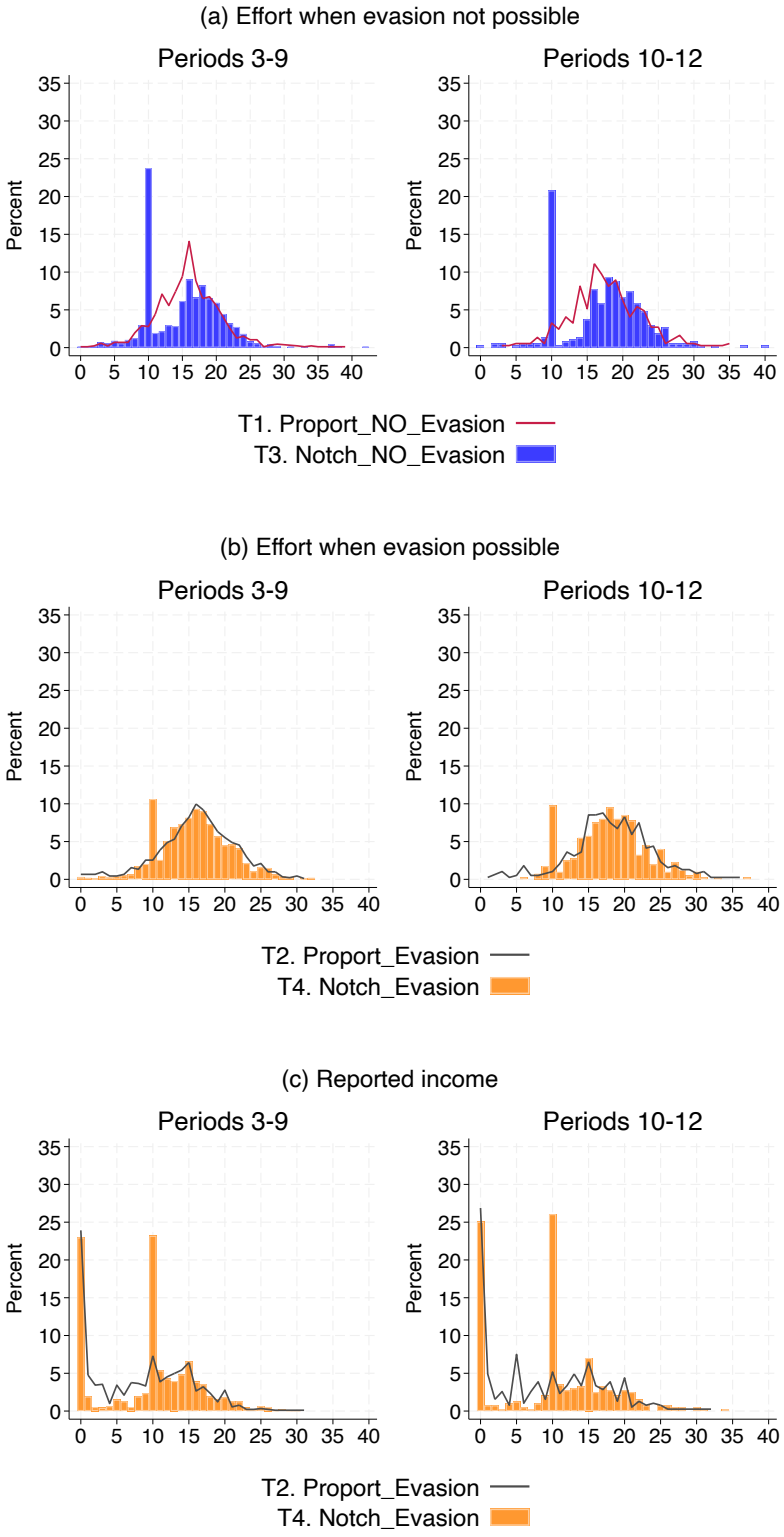
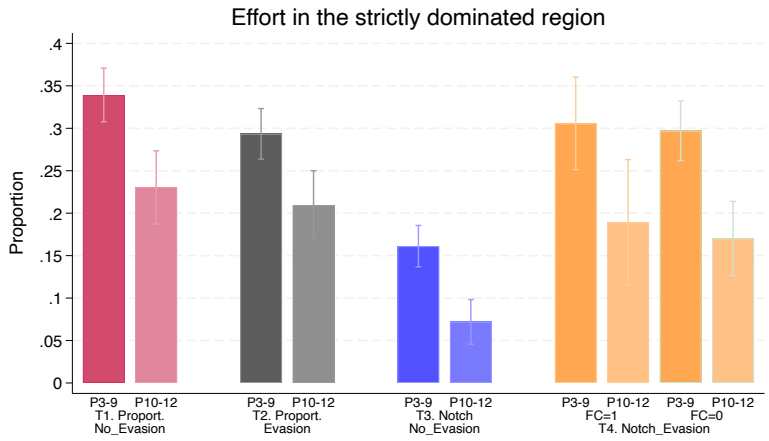


Figure 10: Effort between (10-15] sliders (strictly dominated region) - Periods 3-9 and 10-12.



R6. *Bunching of efforts and of reported incomes are stable over time. Optimization frictions reduce substantially when evasion is not possible. They tend to be more persistent when evasion is permitted.*

5 Concluding discussion

By disentangling the effects of real and reported income responses in tax notch systems, this experimental study contributes to a refined understanding of taxpayer behavior when facing notched fiscal policies. In these conclusions, we summarise our main findings and relate them to questions relevant to both researchers modeling responses to tax notches and policymakers assessing their potential consequences.

In a theoretical model of taxpayer behavior, two main questions arise regarding the effects of a tax notch system: its impact on labor supply and output, and its influence on taxpayers' decisions to evade taxes. Our experiment provides answers to both. In the treatment where underreporting was not possible (treatment 3), participants responded to the tax notch mainly by reducing their labour supply and bunching their effort at the tax threshold to avoid the dominated area created by the notch. When underreporting is possible (treatment 4), we observe more complex adjustments, with a large part of the response affecting reported income. However, we do not observe an increase in tax evasion, as is sometimes feared in policy debates. Instead, we find that subjects tend to reduce evasion, either by reducing effort and opting for full compliance at the notch, or by maintaining effort above the threshold but increasing reported income at the notch

from below the threshold.

These effects are rarely highlighted by the empirical literature. This literature investigates tax notch policies in complex real world tax systems. Nevertheless, the literature faces inherent limitations in fully distinguishing between real and evasion responses and in isolating pure notch effects. Field studies also rely on a procedure to estimate income elasticities from the bunching mass at the notch threshold, assuming it equals the missing mass above it. Our experiment confirms the validity of this assumption when evasion is not possible (as was the case for the original method). However, the assumption becomes more difficult when evasion is possible. This is because, when evasion is possible, both theory and our experiment confirm that a universally dominated area valid for all taxpayers does not exist. Instead, only taxpayers who fully comply should empty the dominated region, while bunching of tax evaders can originate from different potential areas, including areas of potential missing mass below the threshold.

The experiment provides new results on optimization frictions occurring in notch tax systems when individuals fail to empty the region of dominated choices (Kleven and Waseem, 2013; Tazhitdinova, 2020; Adam et al., 2021). As in Gibson et al. (2019), the frictions we document are attributable to cognitive and behavioral effects. We also find that experience reduces these frictions, primarily when labor decisions don't involve tax evasion. Interestingly, when evasion is an option, frictions become more widespread and persistent, even if taxpayers choose not to underreport. Moreover, only taxpayers determined to comply fully should empty the dominated region, whereas tax evaders may still optimally choose to underreport in the region.

These findings can also be relevant when considering the long-run impact of tax notches. When frictions vanish, unresponsive behavior in the dominated area can be added to observed responses to obtain long-run elasticities (e.g., Kleven and Waseem, 2013; Tazhitdinova, 2020). However, since some unresponsive behaviours associated with tax evasion are not frictions or are inherently persistent, this might be taken into account in the assessments of long-run responses to notches when tax evasion is possible.

Notches are characterised by inducing large responses from small changes near the threshold. Our experiment shows that heterogeneity can lead to relevant differences in these responses. We have found that the possibility/non possibility of underreporting mitigated or amplified, respectively, gender differences in performance, with women showing the highest average effort when underreporting was possible. While suggestive, we cannot confirm if this result is related to gender specific effects (similarly to other contexts of choice under pressure; Shurchkov, 2012), or to the interaction between gender with other behavioral attitudes (e.g., reference-dependent effects in notch tax systems;

Kleven, 2016). Future research could further investigate the relationship between individuals' responses to tax notches and their ability and confidence to move beyond dominated regions, including from a gender perspective.

With the qualifications for optimization frictions and heterogeneity, we nevertheless consider the results of our experiment aligning well with the theory of labor supply and tax reporting decisions with quasi-linear preferences and underreporting costs, which is widely used in empirical analyses. This consistency is strengthened by the reliability of the causal inferences in the experiment, achieved by following consolidated practices to design lab experiments in general and tax evasion investigations in particular (Alm and Malézieux, 2021). Internal validity is a crucial prerequisite for considering the potential external validity of the experimental findings (List, 2020).

In this regard, standard caveats that can apply to our experiments include concerns about student sample, small incentives, and artificial settings. This can also pertain to the slider task, which differs from traditional labor supply decisions. However, established arguments, like those made by Doerrenberg and Duncan (2014) in a previous labour supply experiment with tax evasion, mitigate these concerns.[24] Moreover notches in real-world tax systems occur typically as income or other output-related thresholds that taxpayers seek to achieve or miss over a given period, often with some unpredictability, analogously to the thresholds in our lab experiments.

Nonetheless, policy implications from lab experiments always demand extra attention. While the observed reduction in tax evasion might support one argument for tax notch systems, some limitations for this conclusion should be considered. First, our results show that the reduction in tax evasion does not apply to those taxpayers who chose to cheat completely. This finding, which is also consistent with the model with underreporting costs, highlights a potential inherent limit of tax notch policies in combating 'tax ghosts'. Secondly, as previously remarked, our experiment isolated the effect of tax notch systems on tax pressure but did not consider other features of tax systems that can interact with tax notches (Kanbur and Keen, 2014). These primarily include simplified accounting procedures, such as exemptions from accounting obligations, which are often intertwined with notch thresholds in the policies in the field. While sharing the same objective of reducing the tax burden on taxpayers, these simplifications should not compromise enforcement policies and should avoid stimulating an independent in-

[24]These arguments include accumulated evidence showing that higher stakes or non-student populations don't substantially alter experimental outcomes. Moreover, real-effort choices in lab experiments have several features similar to actual labor supply, including comparable demanding and possibly unpleasant aspects (Charness and Kuhn, 2011). More generally, Falk and Heckman (2009) provide a classical and comprehensive discussion on lab experiments as source of knowledge in social science.

crease in tax evasion (Aghion et al., 2024; Di Marzio et al., 2024). Finally, tax equity and other economic and ethical implications of tax notch regimes might pose a concern. As the preferential regimes in the real world derogate from standard tax regimes and are often only accessible to certain categories of taxpayers (e.g., small firms and the self-employed, but generally not to taxpayers under third-party reporting), questions of vertical and horizontal inequalities arise. Even recognizing the potential arguments for these schemes in contexts where tax evasion is a major problem and enforcement policies face constraints (e.g., audit costs and natural limits to the level of fines), significant questions could remain about their long-term effectiveness and the ethical implications of a state that appears to accommodate tax evasion. These broader issues, which have remained beyond our experiment, also require careful consideration for policy implementation.

References

Adam, S., J. Browne, D. Phillips, and B. Roantree (2021). Frictions and taxpayer responses: Evidence from bunching at personal tax thresholds. *International Tax and Public Finance 28*(3), 612–653.

Aghion, P., M. Gravoueille, M. Lequien, and S. Stantcheva (2024). Tax simplicity or simplicity of evasion? Evidence from self-employment taxes in France. Centre for Economic Performance, LSE, Discussion Paper No. 1999.

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

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies. *International Tax and Public Finance 19*, 54–77.

Alm, J. and A. Malézieux (2021). 40 years of tax evasion games: A meta-analysis. *Experimental Economics 24*(3), 699–750.

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

Almunia, M. and D. Lopez-Rodriguez (2018). Under the radar: The effects of monitoring firms on tax compliance. *American Economic Journal: Economic Policy 10*(1), 1–38.

Andreoni, J., B. Erard, and J. Feinstein (1998). Tax compliance. *Journal of Economic Literature 36*(2), 818–860.

Bachas, P. and M. Soto (2021). Corporate taxation under weak enforcement. *American Economic Journal: Economic Policy 13*(4), 36–71.

Bastani, S. and H. Selin (2014). Bunching and non-bunching at kink points of the Swedish tax schedule. *Journal of Public Economics 109*, 36–49.

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

Best, M. C. and H. J. Kleven (2018). Housing market responses to transaction taxes: Evidence from notches and stimulus in the UK. *Review of Economic Studies 85*(1), 157–193.

Blinder, A. S. and H. S. Rosen (1985). Notches. *American Economic Review 75*(4), 736–747.

Buser, T., R. Van Veldhuizen, and Y. Zhong (2025). Time pressure preferences. *Management Science 71*(3), 1909–1924.

Carrillo, P., D. Donaldson, D. Pomeranz, and M. Singhal (2023). Ghosting the tax authority: fake firms and tax fraud in ecuador. *American Economic Review: Insights 5*(4), 427–444.

Carrillo, P., D. Pomeranz, and M. Singhal (2017). Dodging the taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics 9*(2), 144–164.

Charness, G. and P. Kuhn (2011). Chapter 3 - lab labor: What can labor economists learn from the lab? Volume 4 of *Handbook of Labor Economics*, pp. 229–330. Elsevier.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annuual Review of Econonomics 1*(1), 451–488.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records. *Quarterly Journal of Economics 126*(2), 749–804.

Chetty, R., J. N. Friedman, and E. Saez (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the eitc on earnings. *American Economic Review 103*(7), 2683–2721.

Collins, J. H. and R. D. Plumlee (1991). The taxpayer's labor and reporting decision: The effect of audit schemes. *Accounting Review*, 559–576.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic literature 47*(2), 448–474.

Di Marzio, I., S. Mocetti, and E. Rubolino (2024). Market externalities of tax evasion. Temi di Discussione No. 1467, Banca d'Italia.

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

Fajgelbaum, P. D. and A. Khandelwal (2024). The value of de minimis imports. NBER Working Paper Series No. 32607.

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

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Friedland, N., S. Maital, and A. Rutenberg (1978). A simulation study of income tax evasion. *Journal of Public Economics 10*(1), 107–116.

Galasso, V. and P. Profeta (2024). Gender differences in math tests: The role of time pressure. *Economic Journal 134*(664), 3461–3475.

Garbinti, B., J. Goupille-Lebret, M. Munoz, S. Stantcheva, and G. Zucman (2023). Tax design, information, and elasticities: Evidence from the French wealth tax. NBER Working Paper Series No. 31333.

Garicano, L., C. Lelarge, and J. Van Reenen (2016). Firm size distortions and the productivity distribution: Evidence from France. *American Economic Review 106*(11), 3439–3479.

Gibson, J., D. A. Norton, and R. A. White (2019). The backward hustle: An experimental investigation of tax code notches and labor supply. *Journal of Economic Behavior and Organization 166*, 432–445.

Gill, D. and V. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review 102*(1), 469–503.

Gill, D. and V. Prowse (2019). Measuring costly effort using the slider task. *Journal of Behavioral and Experimental Finance 21*, 1–9.

Gourio, F. and N. Roys (2014). Size-dependent regulations, firm size distribution, and reallocation. *Quantitative Economics 5*(2), 377–416.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association 1*(1), 114–125.

Hargaden, E. P. (2020). Taxpayer responses in good times and bad. *Journal of Economic Behavior & Organization 176*, 653–690.

Hiraiwa, T., M. Lipsitz, and E. Starr (2024). Do firms value court enforceability of noncompete agreements? A revealed preference approach. *Review of Economics and Statistics*, 1–47.

Houde, S. (2018). How consumers respond to product certification and the value of energy information. *The RAND Journal of Economics 49*(2), 453–477.

Hungerman, D. (2023). Tax evasion, efficiency, and bunching in the presence of enforcement notches. *International Tax and Public Finance 30*(1), 43–68.

Kanbur, R. and M. Keen (2014). Thresholds, informality, and partitions of compliance. *International Tax and Public Finance 21*, 536–559.

Kaplow, L. (2017). Optimal regulation with exemptions and corrective taxes. Harvard Law School John M. Olin Center Discussion Paper No. 934.

Kastlunger, B., S. G. Dressler, E. Kirchler, L. Mittone, and M. Voracek (2010). Sex differences in tax compliance: Differentiating between demographic sex, gender-role orientation, and prenatal masculinization (2d: 4d). *Journal of Economic Psychology 31*(4), 542–552.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics 8*(1), 435–464.

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

Kleven, H. J. and M. Waseem (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *Quarterly Journal of Economics 128*(2), 669–723.

Le Maire, D. and B. Schjerning (2013). Tax bunching, income shifting and self-employment. *Journal of Public Economics 107*, 1–18.

List, J. A. (2020). Non est disputandum de generalizability? a glimpse into the external validity trial. NBER Working Paper Series No. 27535.

Liu, L., B. Lockwood, M. Almunia, and E. H. Tam (2021). Vat notches, voluntary registration, and bunching: Theory and UK evidence. *Review of Economics and Statistics 103*(1), 151–164.

Lobel, F., T. Scot, and P. Zúniga (2024). Corporate taxation and evasion responses: Evidence from a minimum tax in Honduras. *American Economic Journal: Economic Policy 16*(1), 482–517.

Lockwood, B. (2020). Malas notches. *International Tax and Public Finance 27*, 779–804.

Miller, H., T. Pope, and K. Smith (2024). Intertemporal income shifting and the taxation of business owner-managers. *Review of Economics and Statistics 106*(1), 184–201.

Muthitacharoen, A., W. Wanichthaworn, and T. Burong (2021). Vat threshold and small business behavior: Evidence from Thai tax returns. *International Tax and Public Finance 28*(5), 1242–1275.

OECD (2023). The design of presumptive tax regimes. *Documents de Travail de l'OCDE sur la Fiscalité* (No. 59).

Pántya, J., J. Kovács, C. Kogler, and E. Kirchler (2016). Work performance and tax compliance in flat and progressive tax systems. *Journal of Economic Psychology 56*, 262–273.

Pencavel, J. H. (1979). A note on income tax evasion, labor supply, and nonlinear tax schedules. *Journal of Public Economics 12*(1), 115–124.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy 2*(3), 180–212.

Sallee, J. M. (2014). Rational inattention and energy efficiency. *The Journal of Law and Economics 57*(3), 781–820.

Shurchkov, O. (2012). Under pressure: gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association 10*(5), 1189–1213.

Slemrod, J. (2013). Buenas notches: Lines and notches in tax system design. *eJournalofTax Research 11*, 259.

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

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

Tazhitdinova, A. (2020). Do only tax incentives matter? Labor supply and demand responses to an unusually large and salient tax break. *Journal of Public Economics 184*, 104162.

Traxler, C., F. G. Westermaier, and A. Wohlschlegel (2018). Bunching on the Autobahn? Speeding responses to a notched penalty scheme. *Journal of Public Economics 157*, 78–94.

Webley, P. (1987). Audit probabilities and tax evasion in a business simulation. *Economics Letters 25*(3), 267–270.

Yitzhaki, S. (1974). A note on income tax evasion: A theoretical approach. *Journal of Public Economics 3*(2), 2011–202.

Appendix A Experimental instructions

Instructions for treatment 4 - Translated from Italian. The differences with the other treatments are as follows: the instructions for treatment 3 did not include the part with the possibility to underreport; the instructions for treatment 2 did not include the part on the threshold and the dual tax regimes below and above the threshold; the instructions for treatment 1 did not include either the part with the possibility to underreport or the part on the threshold and the dual tax regimes.

An Experiment on Economic Decision-Making

Welcome and thank you for participating in this experiment. The experiment will last approximately 90 minutes. During the experiment, you will have the opportunity to earn money based on the decisions you make. In addition, you will receive 3 Euro as a participation fee. Your total earnings will be paid to you in cash, individually and privately, at the end of the experiment.

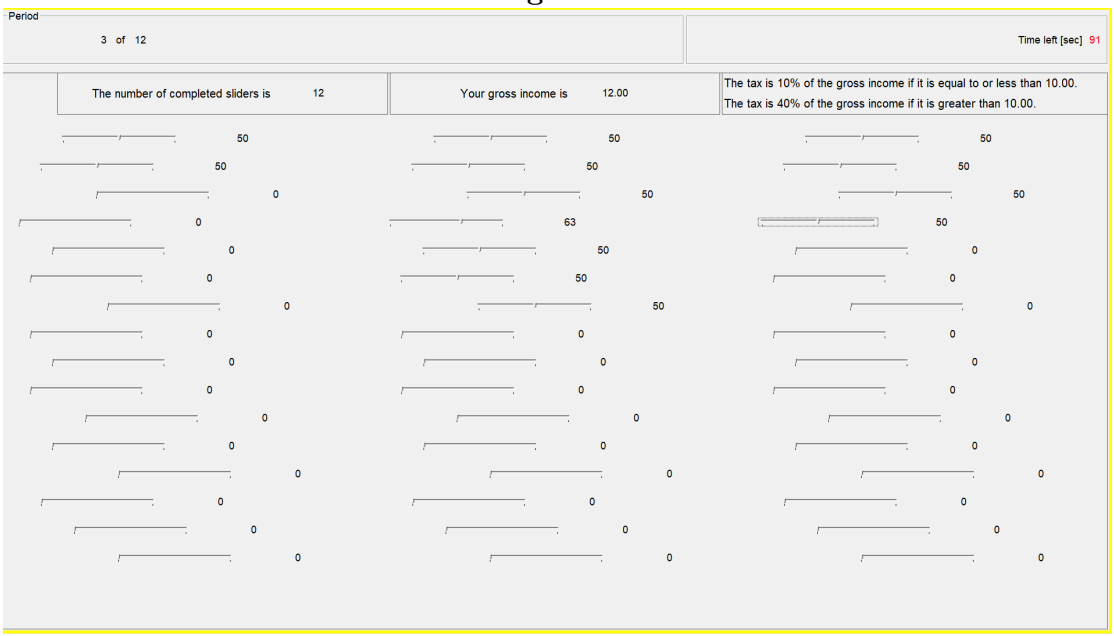
Your decisions and earnings will remain confidential. None of the participants will be informed of the decisions and earnings of other participants, either during the experiment or afterwards. The experimenter will not be able to link your identity to your choices under any circumstances.
An experimenter will now read the instructions aloud. At the end of these instructions, you will be able to ask questions. You will then be asked to answer a few short questions on the computer to check your understanding of how the experiment works. The experiment will then begin, and decisions will be made on the computer. At the end of the experiment, you will be asked to complete a short questionnaire.
During the experiment, you are not allowed to talk or communicate in any way with the other participants. It is very important that you understand the following instructions in order to maximise your earnings beyond the 3 Euro participation fee. If you have any questions during the experiment or while answering the questions to check your understanding of the experiment, click on the "chat" icon at the top right of the screen and an experimenter will come to your terminal to answer your question.

Structure of the experiment
After you have answered the comprehension questions, the experiment will proceed as follows.
This experiment will involve completing individual tasks over 12 periods. The tasks are "individual" because they will have no effect on the tasks and payments of other participants, nor will the tasks of other participants affect your earnings.
The 12 periods will all have the same structure. The tasks you complete in each period will earn you income for that period. The income for the period is subject to taxation. After completing the task, you will be informed of your gross income for the period, i.e. your income before taxation. At this point, you can choose your declared gross income: the declared gross income may be equal to or less than the actual gross income for the period. You will be taxed on the basis of your declared gross income: the lower your declared gross income, the lower the tax you will have to pay for that period. Please note, however, that there is a 10% chance that your decision will be checked to verify that you have declared all your gross income for the period. If you are checked during a period and your declared gross income is lower than your actual gross income, you will have to pay a penalty in addition to the tax. More details are provided below.
At the end of each period, you will be notified whether an audit has been carried out and, if so, you will be informed of your net income (after taxation and any audit of your declared income) for that period.

Your payment for participating in the experiment, in addition to the attendance allowance, will correspond to the net income you earned in only one of the periods of the experiment. In particular, the first two periods are trial periods. Therefore, your earnings for participating will correspond to the net income for one of the periods from the third to the twelfth. At the end of the experiment, the period that counts for payment will be drawn from an envelope containing 10 tickets numbered from 3 to 12 and the period will be the same for all participants.

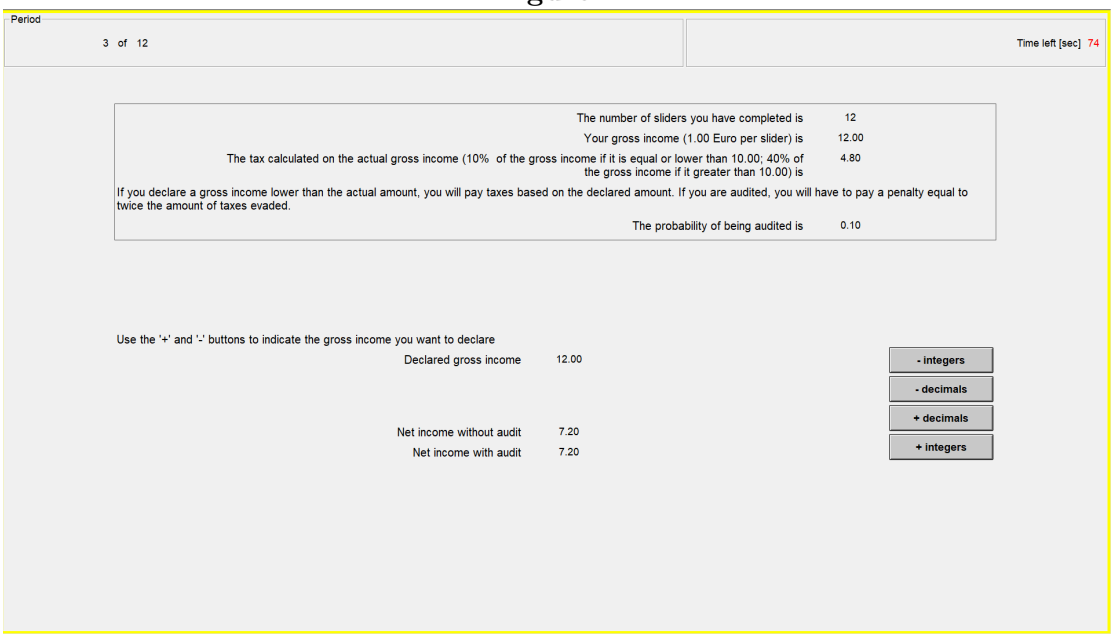
How to earn in the experiment
In each period, you will be shown 48 sliders as shown in Figure 1 below. For each slider, you must try to position the cursor at '50', i.e. exactly in the middle of the slider. The initial position of the cursor is '0'. Moving the cursor with the mouse to the right of the slider will show you the updated position of the cursor [participants could only use the mouse as they had no keyboards available]. In each period, you will have 120 seconds to position the cursors. For each slider positioned correctly, you will earn 1 Euro in gross income. Gross income is subject to taxation. The tax is calculated as follows:
the tax is equal to 10% of the gross income if this is less than or equal to 10 Euro (corresponding to 10 sliders).
the tax is equal to 40% of the gross income if this is greater than 10 Euro (corresponding to 10 sliders).

Figure 1



The example in Figure 1 shows: that the period is 3 of 12; that 12 sliders have been positioned correctly; that the time remaining is 91 seconds; that the gross income obtained up to this point is 12 Euro; that the tax is 10% of the gross income if this is less than or equal to 10 Euro, or 40% of the gross income if it is greater than 10 Euro. At the end of the 120 seconds of each period dedicated to positioning the sliders, a screen like the one in Figure 2 will appear, showing the gross income for the period and asking you to choose the declared gross income.

Figure 2



In the example in Figure 2, the information shows that the screen is for period 3 of 12; that 12 sliders have been completed in the example; that the gross income is 12 Euro; that the tax calculated on this gross income, the effective gross income, would be 4.80 euros (40% of 12 euros, since this gross income is greater than 10 euros); that the probability of being audited is 10%.

Please note, however, that the tax you pay will be calculated based on your declared gross income and not your actual gross income:
tax $= 10\% \times$ declared gross income for a declared gross income less than or equal to 10 Euro.
tax $= 40\% \times$ declared gross income for a declared gross income greater than or equal to 10 Euro.

At this point, you will have 90 seconds to choose the declared gross income that will be taxed. To decide on the declared gross income, you can use the "+" and "-" buttons on the right of the screen. Based on the declared gross income you choose, the net income will be shown on the same screen if you are not audited, and the net income in the event of an audit.

Your net income for the period will be calculated as follows.

If the declared gross income is equal to the actual gross income for the period and this gross income is less than or equal to 10 Euro:
net income = actual gross income − (10%× actual gross income)

If the declared gross income is equal to the actual gross income for the period and this gross income is greater than 10 Euro:
net income = actual gross income − (40%×actual gross income)

If the declared gross income is lower than the actual gross income, the net income depends on the audit.
If the declared gross income is less than or equal to 10 Euro:
net income without audit = actual gross income − declared gross income×10%
net income with audit = actual gross income − declared gross income×10% − penalty

If the declared gross income is greater than 10 Euro:
net income without audit = actual gross income − declared gross income×40%
net income with audit = actual gross income − declared gross income×40% − penalty

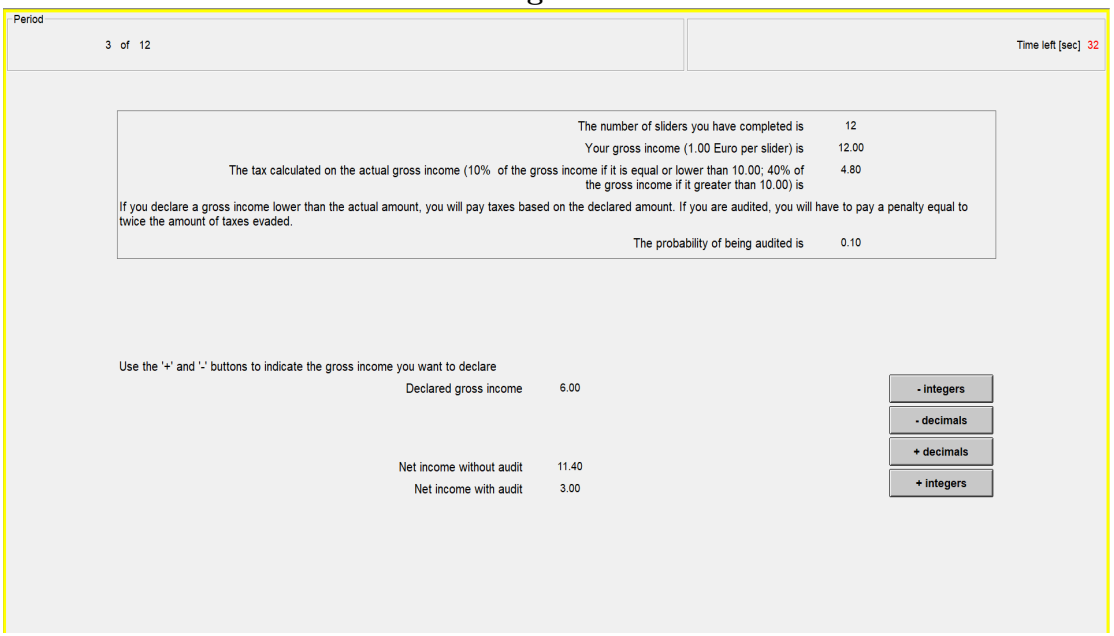
The penalty you will have to pay in the event of an audit will be equal to twice the amount of tax not paid, equal to the difference between the tax due on the actual gross income (10%×actual gross income if the actual gross income is less than or equal to 10 Euro, 40%× if the actual gross income is greater than 10 Euro) and the tax paid on the basis of the declared gross income (10%×declared gross income if the declared gross income is less than or equal to 10 Euro, 40%×declared gross income if the declared gross income is greater than 10 Euro).
The table in the appendix shows, for each number of sliders completed: the gross income, the tax equal to 10% or 40% of the gross income depending on whether it is less than or greater than 10 Euro (corresponding to 10 sliders completed), and the net income when the declared gross income is equal to the actual gross income. The table also highlights that, given the different tax levels for gross income above or below the 10 Euro threshold

corresponding to 10 sliders, the net income between 11 and 15 sliders is lower than that obtained with 10 sliders, when the actual gross income is equal to the declared gross income.

Initially, on the screen that appears after completing the activity with the sliders, the gross income declared will be equal to the actual income, therefore with a net income with audit and a net income without audit that correspond to that reported in the table in the appendix for the number of sliders completed. In the example in Figure 2, the gross income declared is equal to the actual income, 12 Euro. The net income without verification is 7.20 Euro, obtained by subtracting the tax of 4.80 Euro (which is 40%×12 Euro, since the actual gross income is greater than 10 Euro), and is equal to the net income in the case of audit because all gross income has been declared. This net income corresponds to that shown in the table in the appendix for 12 sliders completed. The time remaining is 74 seconds.

In the example in Figure 3, the declared gross income is 6 Euro, which is less than the actual income of 12 Euro. The tax you would pay on this declared gross income would be 0.60 Euro (= 10%×6 Euro, since the declared gross income is less than 10 Euro). The net income without audit would therefore be 11.40 Euro (=12 Euro − 0.60 Euro). The net income in the event of an audit would instead be 3.00 Euro (=12 Euro − 0.60 Euro − 8.40 Euro), as the penalty you would pay if you were audited would be equal to $2 \times (40\% \times 12 - 10\% \times 6)$ =8.40 Euro. The time remaining is 32 seconds.

Figure 3



During the 90 seconds available, you can change the value of your declared gross income as many times as you wish and see the corresponding net income with and without verification. At the end of the 90 seconds available, the declared gross income will be the last one you entered, and on the next screen, you will be informed whether you have been audit and, consequently, your net income for that period will be displayed. Figure 4 follows the example in Figure 3 and shows the net income if no audit has been carried out. In this case, no penalty is payable, so the net income will be 11.40 Euro (=12 Euro − 0.60 Euro).

Figure 4

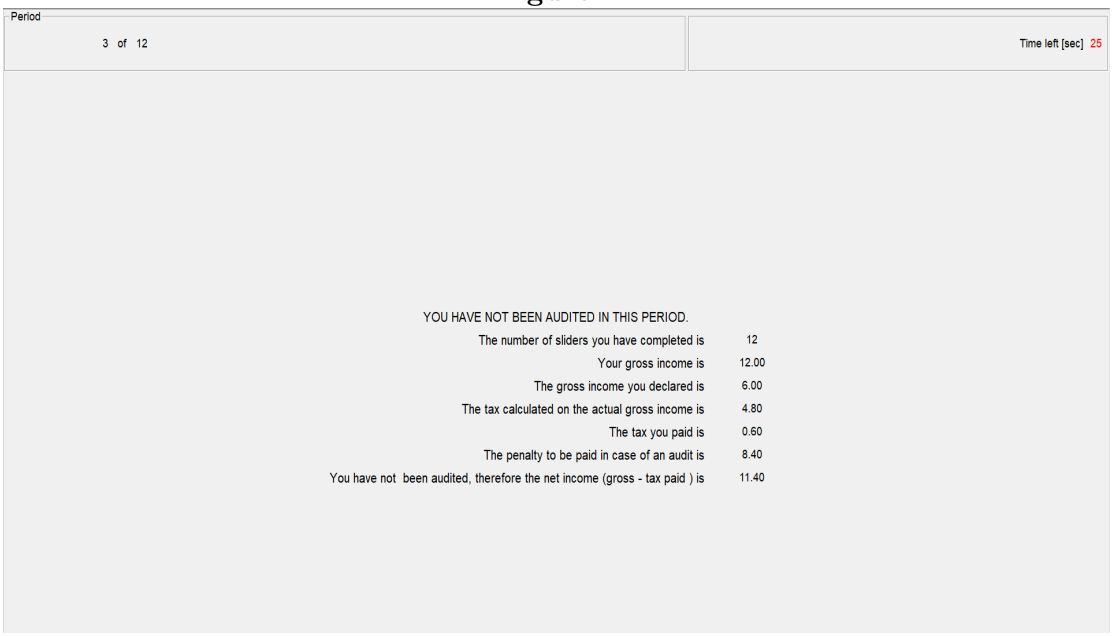
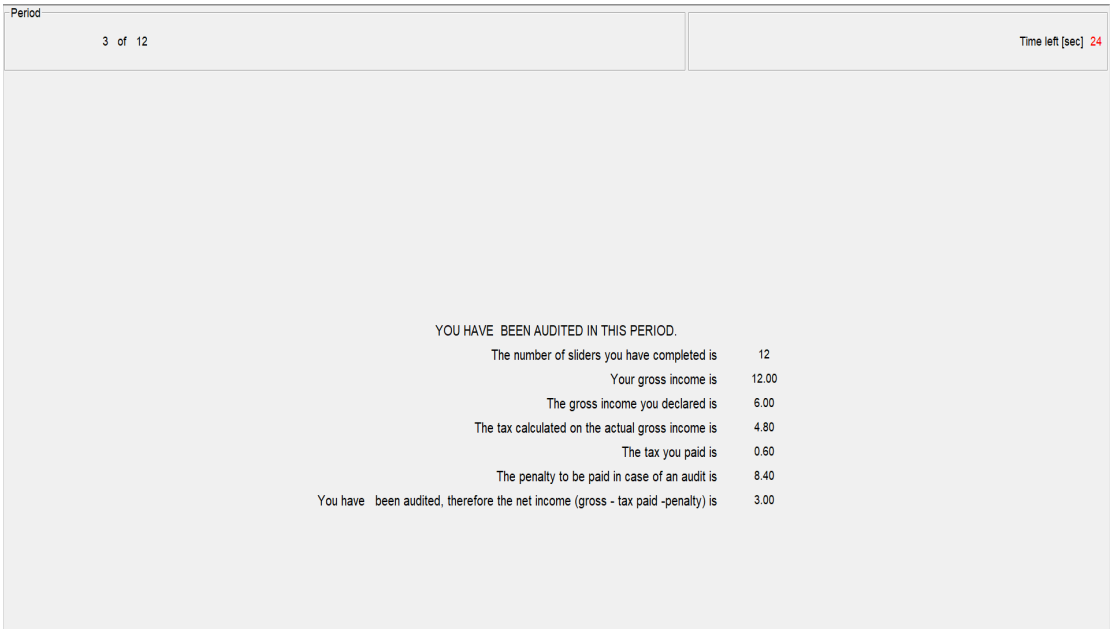


Figure 5 follows the example in Figure 3 and shows the net income in the event of an audit. In this case, the penalty must be paid, so the net income will be 3.00 Euro (=12 Euro − 0.60 Euro − 8.40 Euro).

Figure 5

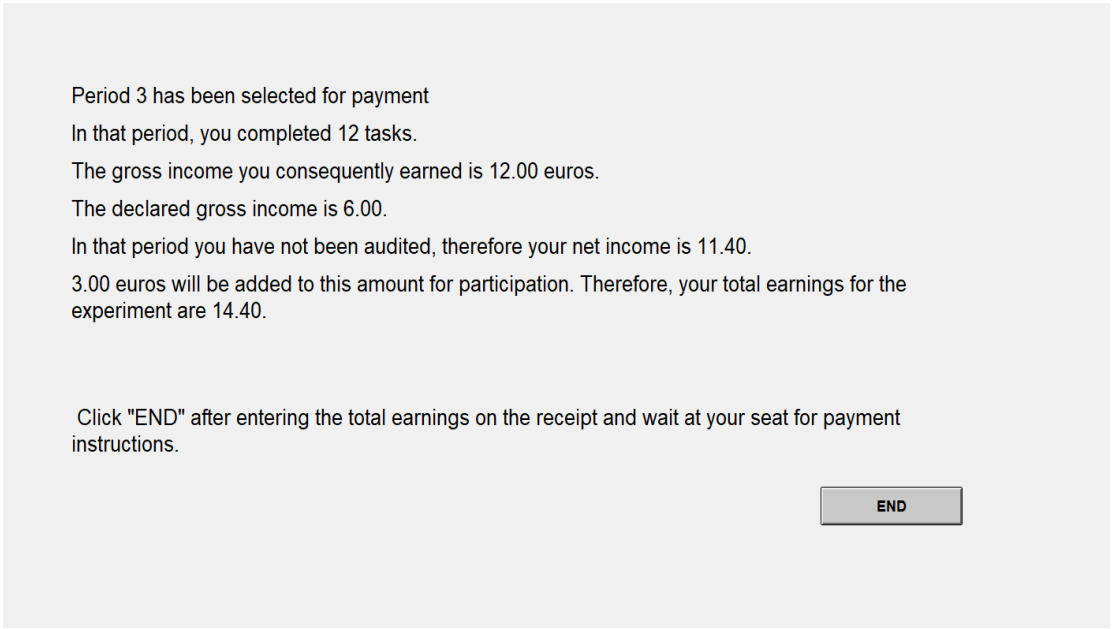


After the information screen, which remains for 30 seconds, you will move on to the next period, where you will again have 120 seconds to position the sliders to obtain the gross income for the new period, which will then be subject to taxation.

End of the experiment

At the end of the 12 periods, a computer screen will inform you that the experiment is over. There will then be a short questionnaire and then the drawing of the period for payment of the experiment. The draw will be made by a participant. Then, the gross income, taxes and net income obtained by each participant in the period drawn will be displayed again on the computer (see Figure 6, for example) and each participant will be paid, including the 3 Euro attendance fee.

Figure 6



Now we invite you to ask for clarification if you have any questions. Then we can move on to the comprehension questions on the computer. Once all participants have answered the questions correctly, the experiment can begin.

APPENDIX – TABLE WITH INCOME AND TAX

Sliders positioned correctly	Gross income	Tax	Net income (gross income - tax)
1	1.00	0.10	0.90
2	2.00	0.20	1.80
3	3.00	0.30	2.70
4	4.00	0.40	3.60
5	5.00	0.50	4.50
6	6.00	0.60	5.40
7	7.00	0.70	6.30
8	8.00	0.80	7.20
9	9.00	0.90	8.10
10	10.00	1.00	9.00
11	11.00	4.40	6.60
12	12.00	4.80	7.20
13	13.00	5.20	7.80
14	14.00	5.60	8.40
15	15.00	6.00	9.00
16	16.00	6.40	9.60
17	17.00	6.80	10.20
18	18.00	7.20	10.80
19	19.00	7.60	11.40
20	20.00	8.00	12.00
21	21.00	8.40	12.60
22	22.00	8.80	13.20
23	23.00	9.20	13.80
24	24.00	9.60	14.40
25	25.00	10.00	15.00
26	26.00	10.40	15.60
27	27.00	10.80	16.20
28	28.00	11.20	16.80
29	29.00	11.60	17.40
30	30.00	12.00	18.00
31	31.00	12.40	18.60
32	32.00	12.80	19.20
33	33.00	13.20	19.80
34	34.00	13.60	20.40
35	35.00	14.00	21.00
36	36.00	14.40	21.60
37	37.00	14.80	22.20
38	38.00	15.20	22.80
39	39.00	15.60	23.40
40	40.00	16.00	24.00
41	41.00	16.40	24.60
42	42.00	16.80	25.20
43	43.00	17.20	25.80
44	44.00	17.60	26.40
45	45.00	18.00	27.00
46	46.00	18.40	27.60
47	47.00	18.80	28.20
48	48.00	19.20	28.80

Pictures of the laboratory

Appendix B Sample participants and model estimates

Table B.1: Summary statistics of the sample participants.

	T1. Proport NO Evasion	T2. Proport Evasion	T3. Notch NO Evasion	T4. Notch Evasion	Total
N	123 (24.1%)	129 (25.3%)	125 (24.5%)	133 (26.1%)	510 (100.0%)
Gender					
Female	64 (52.0%)	64 (49.6%)	65 (52.0%)	73 (54.9%)	266 (52.2%)
Male	57 (46.3%)	63 (48.8%)	58 (46.4%)	60 (45.1%)	238 (46.7%)
Prefer not to say	2 (1.6%)	2 (1.6%)	2 (1.6%)	0 (0.0%)	6 (1.2%)
Age					
1: 18 y.o.	0 (0.0%)	1 (0.8%)	3 (2.4%)	0 (0.0%)	4 (0.8%)
2: 19 y.o.	27 (22.0%)	22 (17.1%)	21 (16.8%)	7 (5.3%)	77 (15.1%)
3: 20 y.o.	20 (16.3%)	24 (18.6%)	21 (16.8%)	19 (14.3%)	84 (16.5%)
4: 21 y.o.	26 (21.1%)	16 (12.4%)	17 (13.6%)	24 (18.0%)	83 (16.3%)
5: 22 y.o.	9 (7.3%)	13 (10.1%)	10 (8.0%)	19 (14.3%)	51 (10.0%)
6: 23 y.o.	12 (9.8%)	19 (14.7%)	25 (20.0%)	23 (17.3%)	79 (15.5%)
7: 24-30 y.o.	26 (21.1%)	32 (24.8%)	25 (20.0%)	32 (24.1%)	115 (22.5%)
8: >30 y.o.	3 (2.4%)	2 (1.6%)	3 (2.4%)	9 (6.8%)	17 (3.3%)
Age Avg. (Sd.)	4.398 (1.908)	4.651 (1.920)	4.600 (1.940)	5.233 (1.708)	4.729 (1.867)
Experience in exp.					
0 exp	30 (24.4%)	27 (20.9%)	30 (24.0%)	26 (19.5%)	113 (22.2%)
1 exp	19 (15.4%)	22 (17.1%)	22 (17.6%)	22 (16.5%)	85 (16.7%)
2 exp	18 (14.6%)	21 (16.3%)	25 (20.0%)	20 (15.0%)	84 (16.5%)
>2 exp	56 (45.5%)	59 (45.7%)	48 (38.4%)	65 (48.9%)	228 (44.7%)
Experience Avg. (Sd.)	1.813 (1.246)	1.868 (1.204)	1.728 (1.203)	1.932 (1.197)	1.837 (1.212)
Study level					
Undergraduate	87 (70.7%)	82 (63.6%)	71 (56.8%)	67 (50.4%)	307 (60.2%)
Postgraduate	28 (22.8%)	39 (30.2%)	50 (40.0%)	61 (45.9%)	178 (34.9%)
Other	8 (6.5%)	8 (6.2%)	4 (3.2%)	5 (3.8%)	25 (4.9%)
Study field					
Economics	49 (39.8%)	49 (38.0%)	40 (32.0%)	34 (25.6%)	172 (33.7%)
Management	25 (20.3%)	25 (19.4%)	30 (24.0%)	37 (27.8%)	117 (22.9%)
Lang. & Human.	49 (39.8%)	55 (42.6%)	55 (44.0%)	62 (46.6%)	221 (43.3%)
Risk	5.520 (1.993)	5.682 (2.246)	5.656 (1.914)	5.368 (2.578)	5.555 (2.203)

Notes. This table reports summary statistics for the sample participants based on the self-reported questionnaire at the end of the experiment. 'Risk' is the subjects' self-rated propensity to take risky behaviour on a ten-point scale.

Table B.2: Regressions of choice variables on treatments and controls (all coefficients).

	(1) Effort	(2) Reported income	(3) Under-reporting rate	(4) Effort	(5) Reported income	(6) Under-reporting rate
T2. Proport_Evasion	0.346			0.764		
	(0.567)			(0.738)		
T3. Notch_NO_Evasion	-0.606			0.165		
	(0.583)			(0.868)		
T4. Notch_Evasion	-0.001	1.471*	-0.105**	-1.070	0.104	-0.072
	(0.530)	(0.729)	(0.041)	(0.873)	(1.135)	(0.057)
T2. Proport_Evasion × Female=1				-1.087		
				(1.046)		
T3. Notch_NO_Evasion × Female=1				-1.532		
				(1.103)		
T4. Notch_Evasion × Female=1				2.111*	2.636	-0.038
				(1.069)	(1.452)	(0.077)
Female=1				-2.787***	-0.592	-0.091
				(0.783)	(1.082)	(0.058)
Age				0.013	-0.296	0.002
				(0.166)	(0.282)	(0.016)
Postgraduate				-0.120	-0.714	0.069
				(0.615)	(0.978)	(0.056)
Non-student				-0.472	0.606	-0.028
				(0.965)	(1.635)	(0.087)
Management				-0.518	0.494	-0.048
				(0.533)	(1.048)	(0.054)
Lang.& Human.				0.020	1.390	-0.120*
				(0.488)	(0.864)	(0.047)
Experience 1				0.894	-0.840	0.036
				(0.566)	(0.985)	(0.055)
Experience 2				1.413*	-0.726	0.072
				(0.654)	(1.284)	(0.064)
Experience 3				0.858	-0.175	0.003
				(0.473)	(0.956)	(0.051)
Risk				0.085	-0.459**	0.032***
				(0.092)	(0.156)	(0.008)
Trend (linear)	0.440***	0.117**	0.009***	0.440***	0.117**	0.009***
	(0.017)	(0.039)	(0.002)	(0.017)	(0.039)	(0.002)
Constant	14.000***	7.600***	0.442***	14.341***	11.711***	0.316**
	(0.386)	(0.526)	(0.032)	(0.991)	(1.900)	(0.098)
Obs.	5100	2620	2620	5100	2620	2620
Clust.	510	262	262	510	262	262
F	174.27	6.59	11.50	48.08	2.91	5.60
Prob >F	0.0000	0.0016	0.0000	0.0000	0.0006	0.0000
R-squared	0.0610	0.0132	0.0219	0.1533	0.0655	0.1208

Notes. This table reports full estimates, including all control coefficients (not reported in Table 2 of Section 4.1) from regressions of choice variables over periods 3-12 (payment periods). Standard errors are clustered at the subject level and reported in parentheses. For regressions in columns (1) and (4), the baseline is T1. Proport_NO_Evasion; for regressions in columns (2)-(3) and (5)-(6), the baseline is T2. Proport_Evasion. Statistical significance is indicated as: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table B.3: Estimated coefficients of effort choices across treatments (Multinomial Logit).

	(1) Eff.< 10	(2) Str. domin. Eff. $(10-15]$	(3) Eff. $(15-18]$	(4) Eff. > 18
2. Proport_Evasion	0.229	0.134	0.469	0.807
	(0.882)	(0.568)	(0.646)	(0.671)
3. Notch_NO_Evasion	-2.564***	-3.327***	-2.375***	-2.047***
	(0.709)	(0.418)	(0.498)	(0.544)
4. Notch_Evasion	-1.259	-1.849***	-1.924***	-1.909**
	(0.690)	(0.466)	(0.543)	(0.584)
Female=1	0.228	-0.395	-0.770	-1.560**
	(0.627)	(0.456)	(0.549)	(0.601)
2. Proport_Evasion×Female=1	0.410	0.150	-0.413	-0.590
	(0.979)	(0.690)	(0.795)	(0.858)
3. Notch_NO_Evasion×Female=1	0.616	0.551	-0.050	-0.085
	(0.817)	(0.542)	(0.634)	(0.711)
4. Notch_Evasion×Female=1	-0.492	0.665	0.784	1.367
	(0.825)	(0.596)	(0.692)	(0.756)
Risk	0.016	0.101*	0.100*	0.081
	(0.068)	(0.045)	(0.051)	(0.055)
Trend (linear)	-0.136***	-0.058**	0.040*	0.161***
	(0.027)	(0.021)	(0.020)	(0.018)
Constant	1.685*	2.742***	2.179**	1.890*
	(0.757)	(0.617)	(0.696)	(0.755)
Controls	Yes	Yes	Yes	Yes
Obs.	5100			
Clust.	510			
Chi-2	674.86			
Prob >Chi-2	0.0000			
Pseudo R-2	0.0895			

Notes. This table reports the coefficients estimated from the Multinomial Logit model for the choices of effort across treatments. The estimates are used to obtain the predicted probabilities and marginal effects for the whole sample in Table 3 of Section 4.2, and those by gender in Tables C.1 and C.2 of Appendix C. The base category for estimation is effort at 10 (bunched). The baseline is T1.Proport_NO_Evasion. Controls are those reported in Table B.2. Standard errors are clustered at the subject level and reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.4: Estimated coefficients of bunching and strictly dominated effort choices in treatment 4 controlling for full compliance (FC) (Logit Models).

	(1) Bunching Eff. = 10	(2) Str. domin. Eff. $(10-15]$
FC=1	13.219***	0.880
	(7.507)	(0.290)
Female=1	0.950	1.265
	(0.583)	(0.359)
FC=1 × Female=1	0.522	1.344
	(0.377)	(0.559)
Risk	0.968	1.056
	(0.064)	(0.046)
Trend (linear)	0.963	0.889***
	(0.038)	(0.023)
Controls	Yes	Yes
Obs.	1330	1330
Clust.	133	133
Chi-2	51.17	49.95
Prob >Chi-2	0.0000	0.0000
Log pseudolikelihood	-355.2569	-734.0441

Notes. This table reports the coefficients estimated from the logit models for bunching and strictly dominated effort choices in treatment 4. The estimates are used to obtain the predicted probabilities and marginal effects for the whole sample in Table 4 of Section 4.3, and those by gender in Table C.3 of Appendix C. Controls are those reported in Table B.2. Standard errors are clustered at the subject level and reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.5: Estimated coefficients of joint reported income and effort choices in treatments 2 and 4 (Multinomial Logit).

	(1) Eff.<10 FC	(2) Eff.<10 Ev	(3) Eff.=10 Ev	(4) Eff.>10 FC	(5) Eff.>10 Rep.<10	(6) Eff.>10 Rep.=10	(7) Eff.>Rep. Rep.>10
T4. Notch_Evasion	-1.342	-3.064**	-1.480	-3.327***	-3.573***	-2.103*	-3.181***
	(0.843)	(0.992)	(1.063)	(0.794)	(0.798)	(0.874)	(0.810)
Female=1	1.465	1.081	1.048	-0.405	-0.740	-0.195	-0.481
	(0.840)	(0.912)	(1.107)	(0.818)	(0.833)	(0.912)	(0.862)
T4. Notch_Evasion × Female=1	-1.989	-0.783	-0.821	1.314	0.476	0.723	0.953
	(1.115)	(1.127)	(1.307)	(0.923)	(0.969)	(1.041)	(0.978)
Risk	0.039	0.086	0.151	0.011	0.244**	0.112	0.060
	(0.108)	(0.133)	(0.121)	(0.079)	(0.094)	(0.085)	(0.080)
Trend (linear)	-0.207**	-0.101*	-0.033	0.051	0.061	0.048	0.053
	(0.067)	(0.051)	(0.071)	(0.036)	(0.036)	(0.042)	(0.035)
Constant	0.865	3.359*	2.151	4.379***	4.394**	3.193*	5.439***
	(1.389)	(1.346)	(1.688)	(1.254)	(1.342)	(1.371)	(1.330)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	2620						
Clust.	262						
Chi-2	549.16						
Prob >Chi-2	0.0000						
Pseudo R-2	0.0904						

Notes. This table reports the coefficients estimated from the Multinomial Logit model for the joint reported income and effort choices in treatments 2 and 4. The estimates are used to obtain the predicted probabilities and marginal effects for the whole sample in Table 5 of Section 4.4, and those by gender in Table C.4 of Appendix C. In columns (1) and (4), FC indicates choice with full-compliance; in columns (2) and (3), Ev indicates choice with evasion; in columns (5)-(7) all choice are with evasion for different levels of reported income. The base category for estimation is effort at 10 with full compliance. The baseline is T2. Proport_Evasion. Controls are those reported in Table B.2. Standard errors are clustered at the subject level and reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.6: Estimated coefficients of underreporing rates (URR) in treatments 2 and 4 (Multinomial Logit).

	(1) Full compliance	(2) Medium-large evasion (0.5 <URR< 1)	(3) Total evasion
T4. Notch_Evasion	0.320	-1.102**	0.020
	(0.390)	(0.419)	(0.362)
Female=1	-0.035	-0.191	-0.721
	(0.385)	(0.361)	(0.417)
T4. Notch_Evasion × Female=1	0.033	0.200	-0.402
	(0.487)	(0.576)	(0.580)
Risk	-0.065	0.142*	0.151*
	(0.050)	(0.061)	(0.068)
Trend (linear)	-0.008	0.041	0.049**
	(0.016)	(0.022)	(0.016)
Constant	-1.139	-1.310*	-1.667*
	(0.594)	(0.641)	(0.779)
Controls	Yes	Yes	Yes
Obs.	2620		
Clust.	262		
Chi-2	95.71		
Prob >Chi-2	0.0000		
Pseudo R-2	0.0723		

Notes. This table reports the coefficients estimated from the Multinomial Logit model for the underreporting rates in treatments 2 and 4. The estimates are used to obtain the predicted probabilities and marginal effects for the whole sample in Table 6 of Section 4.5, and those by gender in Tables C.5 of Appendix C. The base category for estimation is medium-small evasion (0 <URR≤ 5). The baseline is T2. Proport_Evasion. Controls are those reported in Table B.2. Standard errors are clustered at the subject level and reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix C Breakdown of estimates by gender

This appendix reports the predicted probabilities and marginal effects separately for females and males, obtained by the models in Appendix B. The models are the same as those used for the predicted probabilities and marginal effects for the whole sample in the paper.

Specifically, Tables C.1 and C.2 show the separate predicted probabilities and marginal effects for females and males, based on the five effort categories of the multinomial logit model in Table B.3. Overall, the predicted probabilities and marginal effects confirm for both genders the result R1 of the text. The main differences between genders highlighted by the marginal effects (Tab. C.2). Females perform worse in the slider task than males in treatments 1 and 2, exhibiting more frequent choices in categories with fewer sliders, including those with 10-15 sliders. Consequently, they appear more challenged by the strictly dominated region and bunch effort in treatment 3 more than males (col. 2). Gender differences tend to disappear in treatment 4, as females exert significantly more effort than in treatment 3 (including above 18 sliders, col. 5). This is likely because the possibility of underreporting mitigates their fear of ending up in the strictly dominated region. Males are less challenged by the strictly dominated region even in treatment 3 and exhibit less bunching. They also tend to reduce effort in the highest effort category of treatment 4 (above 18 sliders, col. 5), particularly compared to treatment 2, where evasion is possible under the proportional system.

Table C.3 shows the breakdown by gender of the predicted probabilities and marginal effects for effort at the threshold and in the strictly dominated area, distinguishing in treatment 4 between full compliance (FC=1) and evasion (FC=0). The estimates for both genders confirm the R2 findings of the text, showing that optimization frictions in the strictly dominated region remain high under full compliance in treatment 4 (as also shown in Figure C.1).

Table C.4 shows, by gender, the predicted probabilities and marginal effects for the eight categories of joint reported income and income decisions estimated by the multinomial logit model in Table B.5. These results confirm the R3 findings of the text for both genders. For both genders in treatment 4, there is a large decrease in effort levels greater than 10 with reported income less than 10 (col. 6), accompanied by a an increase in effort levels at the threshold with full compliance (col. 3) and in effort levels greater than 10 with reported income at 10 (col. 7). Conversely, there is no increase in evasion at the threshold accompanied by a decrease in the categories of effort and reported income above the threshold (col. 5 and col. 8).

Table C.5 reports the predicted probabilities and marginal effects by gender for the

Table C.1: Predicted probabilities of effort choices across treatments, by gender (from Multinomial Logit model in Table B.3).

	(1) Eff.< 10	(2) Bunching Eff. $= 10$	(3) Str. domin. Eff. $(10-15]$	(4) Eff. $(15-18]$	(5) Eff. > 18
Predicted probabilities					
T1. Proport_NO_Evasion × Female=1	0.10	0.04	0.36	0.30	0.20
	(0.026)	(0.010)	(0.034)	(0.029)	(0.034)
T1. Proport_NO_Evasion × Female=0	0.04	0.02	0.25	0.29	0.41
	(0.018)	(0.007)	(0.032)	(0.029)	(0.048)
T2. Proport_Evasion × Female=1	0.15	0.03	0.37	0.25	0.20
	(0.033)	(0.009)	(0.034)	(0.030)	(0.038)
T2. Proport_Evasion × Female=0	0.03	0.01	0.17	0.27	0.52
	(0.016)	(0.005)	(0.024)	(0.025)	(0.043)
T3. Notch_NO_Evasion × Female=1	0.12	0.30	0.18	0.21	0.19
	(0.029)	(0.041)	(0.024)	(0.026)	(0.034)
T3. Notch_NO_Evasion × Female=0	0.03	0.16	0.08	0.25	0.48
	(0.013)	(0.031)	(0.014)	(0.025)	(0.044)
T4. Notch_Evasion × Female=1	0.05	0.10	0.30	0.25	0.30
	(0.012)	(0.022)	(0.029)	(0.025)	(0.039)
T4. Notch_Evasion × Female=0	0.06	0.10	0.23	0.25	0.36
	(0.020)	(0.028)	(0.032)	(0.028)	(0.047)

Notes. Standard errors clustered at the subject level are in parentheses. (As in the tables in the text, statistical significance is not reported for predicted probabilities).

underreporting rates in four classes estimated by the multinomial logit model in Table B.6. These results confirm R4 of the text for both genders, namely that a statistically significant decrease in large-scale evasion occurs in treatment 4 compared to treatment 2, while the choices of total evasion remain unchanged (see also Figure C.2).

Table C.2: Marginal effects of effort choices across treatments, by gender (from Multinomial Logit model in Table B.3 and predicted probabilities in Table C.1).

	(1) Eff.< 10	(2) Bunching Eff. $= 10$	(3) Str. domin. Eff. $(10-15]$	(4) Eff. $(15-18]$	(5) Eff. > 18
Marginal effects Female					
T2 vs. T1	0.047	-0.008	0.014	-0.050	-0.002
	(0.040)	(0.013)	(0.047)	(0.041)	(0.051)
T3 vs. T1	0.015	0.262***	-0.178***	-0.087*	-0.011
	(0.038)	(0.042)	(0.041)	(0.038)	(0.047)
T4 vs. T1	-0.053	0.063**	-0.063	-0.047	0.099
	(0.028)	(0.024)	(0.045)	(0.038)	(0.051)
T4 vs. T2	-0.100**	0.071**	-0.076	0.003	0.102
	(0.035)	(0.023)	(0.045)	(0.039)	(0.054)
T4 vs. T3	-0.067*	-0.199***	0.116**	0.040	0.111*
	(0.031)	(0.048)	(0.037)	(0.036)	(0.051)
Marginal effects Male					
T2 vs. T1	-0.009	-0.007	-0.078*	-0.019	0.113
	(0.024)	(0.008)	(0.040)	(0.038)	(0.063)
T3 vs. T1	-0.010	0.143***	-0.163***	-0.039	0.069
	(0.022)	(0.032)	(0.035)	(0.038)	(0.064)
T4 vs. T1	0.025	0.084**	-0.020	-0.040	-0.048
	(0.027)	(0.029)	(0.046)	(0.040)	(0.068)
T4 vs. T2	0.034	0.091**	0.058	-0.022	-0.161*
	(0.026)	(0.029)	(0.040)	(0.038)	(0.064)
T4 vs. T3	0.035	-0.059	0.143***	-0.001	-0.117
	(0.024)	(0.041)	(0.035)	(0.037)	(0.064)
Marginal effects Female vs. Male					
T1	0.062*	0.021	0.111*	0.011	-0.205***
	(0.032)	(0.013)	(0.047)	(0.041)	(0.059)
T2	0.119***	0.020*	0.203***	-0.021	-0.321***
	(0.035)	(0.010)	(0.042)	(0.040)	(0.058)
T3	0.087**	0.140**	0.096***	-0.037	-0.286***
	(0.033)	(0.052)	(0.028)	(0.036)	(0.056)
T4	-0.015	0.000	0.069	0.004	-0.058
	(0.024)	(0.036)	(0.043)	(0.038)	(0.061)

Notes. Standard errors clustered at the subject level are in parentheses. Statistical significance is indicated as: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

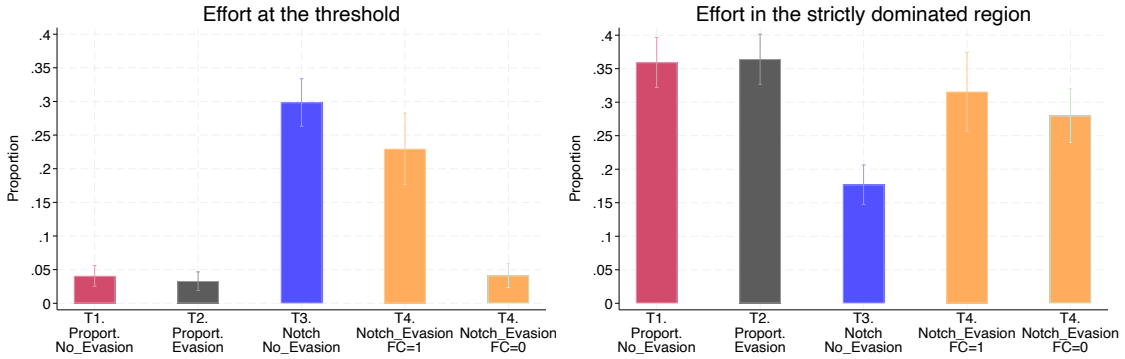
Table C.3: Predicted probabilities and marginal effects of bunching and strictly dominated effort choices in treatment 4, by gender (from Logit models in Table B.4).

	(1) Bunching Eff. = 10	(2) Str. domin. Eff. $(10-15]$
Predicted probabilities		
FC=1×Female=1	0.21	0.31
	(0.042)	(0.048)
FC=0×Female=1	0.04	0.28
	(0.014)	(0.033)
FC=1×Female=0	0.33	0.21
	(0.068)	(0.050)
FC=0×Female=0	0.04	0.23
	(0.017)	(0.037)
Marginal effects		
Female: FC=1 vs. FC=0	0.167***	0.033
	(0.043)	(0.054)
Male: FC=1 vs. FC=0	0.292***	-0.021
	(0.054)	(0.054)
FC=1: Female vs. Male	-0.127	0.098
	(0.078)	(0.071)
FC=0: Female vs. Male	-0.002	0.043
	(0.023)	(0.052)

Notes. Standard errors clustered at the subject level are in parentheses. Stars for statistical significance are only reported for marginal effects: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure C.1: Bunching of effort and emptying of the strictly dominated area, by gender.

(a) - Female



(b) - Male

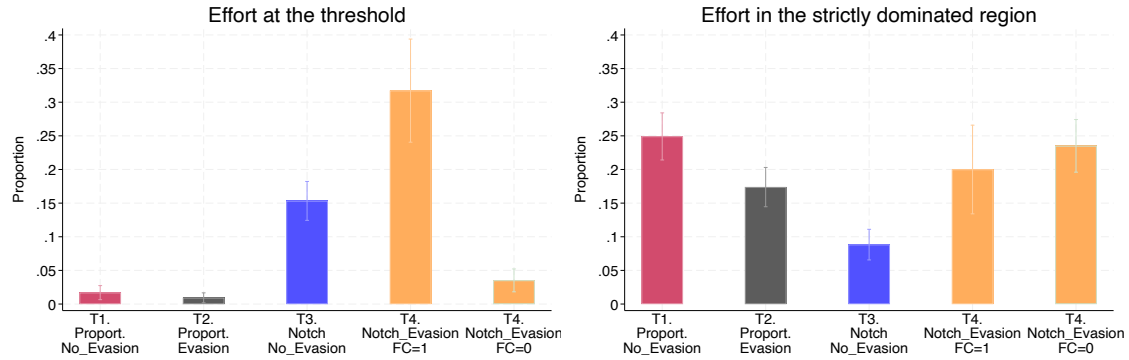


Table C.4: Predicted probabilities and marginal effects of joint reported income and effort choices in treatments 2 and 4 (from Multinomial Logit molde in Table B.5).

	(1) Eff.<10 FC	(2) Eff.<10 Ev	(3) Eff.=10 FC	(4) Eff.=10 Ev	(5) Eff.>10 FC	(6) Eff.>10 Rep.<10	(7) Eff.>10 Rep.=10	(8) Eff.>Rep. Rep.>10
Predicted probabilities								
T2. Proport_Evasion × Female=1	0.04	0.10	0.01	0.02	0.15	0.37	0.06	0.24
	(0.014)	(0.023)	(0.003)	(0.009)	(0.030)	(0.041)	(0.011)	(0.039)
T4. Notch_Evasion × Female=1	0.02	0.03	0.07	0.03	0.22	0.20	0.16	0.28
	(0.007)	(0.008)	(0.017)	(0.011)	(0.035)	(0.037)	(0.029)	(0.038)
T2. Proport_Evasion × Female=0	0.01	0.02	0.00	0.01	0.15	0.50	0.05	0.26
	(0.005)	(0.014)	(0.003)	(0.004)	(0.039)	(0.047)	(0.012)	(0.037)
T4. Notch_Evasion × Female=0	0.04	0.03	0.09	0.03	0.12	0.33	0.12	0.24
	(0.015)	(0.010)	(0.029)	(0.012)	(0.026)	(0.050)	(0.029)	(0.045)
Marginal effects								
Female: T4 vs. T2	-0.027	-0.076**	0.060***	0.003	0.069	-0.176**	0.103**	0.044
	(0.016)	(0.023)	(0.017)	(0.013)	(0.046)	(0.055)	(0.032)	(0.055)
Male: T4 vs. T2	0.033*	0.001	0.085**	0.024	-0.033	-0.170**	0.079*	-0.018
	(0.016)	(0.017)	(0.029)	(0.012)	(0.045)	(0.066)	(0.032)	(0.056)
T2: Female vs. Male	0.037*	0.078***	0.002	0.019*	-0.001	-0.129*	0.011	-0.016
	(0.015)	(0.023)	(0.004)	(0.010)	(0.051)	(0.065)	(0.017)	(0.056)
T4: Female vs. Male	-0.022	0.000	-0.023	-0.002	0.102*	-0.135*	0.034	0.046
	(0.017)	(0.013)	(0.033)	(0.017)	(0.044)	(0.064)	(0.040)	(0.059)

Notes. Standard errors clustered at the subject level are in parentheses. Stars for statistical significance are only reported for marginal effects: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

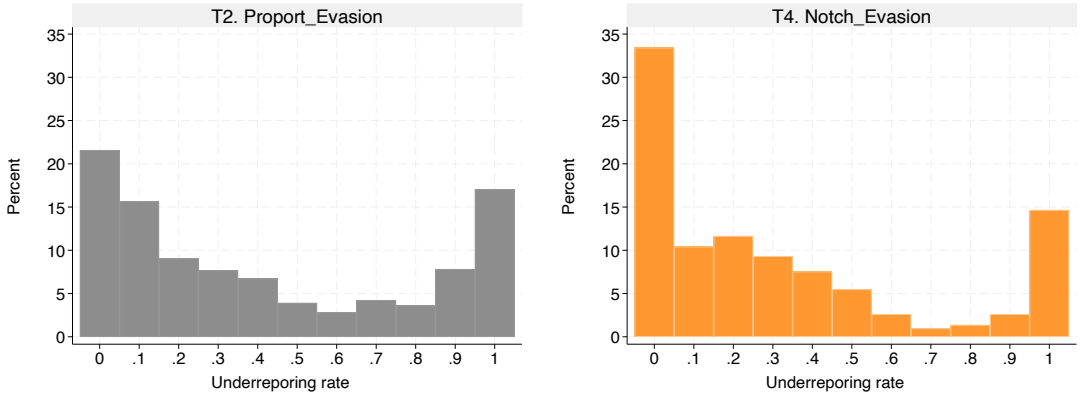
Table C.5: Predicted probabilities and marginal effects of underreporting rates (URR) in treatments 2 and 4, by gender (from Multinomial Logit model in Table B.6).

	(1) Full compliance	(2) Medium-small evasion (0 <URR≤ 0.5)	(3) Medium-large evasion (0.5 <URR< 1)	(4) Total evasion
Predicted probabilities				
T2. Proport_Evasion × Female=1	0.20	0.40	0.21	0.19
	(0.032)	(0.046)	(0.042)	(0.041)
T4. Notch_Evasion × Female=1	0.31	0.44	0.10	0.15
	(0.040)	(0.045)	(0.022)	(0.034)
T2. Proport_Evasion × Female=0	0.17	0.32	0.21	0.30
	(0.040)	(0.039)	(0.034)	(0.048)
T4. Notch_Evasion × Female=0	0.25	0.35	0.08	0.33
	(0.042)	(0.048)	(0.023)	(0.050)
Marginal effects				
Female: T4 vs. T2	0.108*	0.045	-0.113*	-0.039
	(0.051)	(0.064)	(0.048)	(0.052)
Male: T4 vs. T2	0.075	0.023	-0.130**	0.032
	(0.056)	(0.061)	(0.042)	(0.067)
T2: Female vs. Male	0.030	0.075	0.008	-0.113
	(0.053)	(0.063)	(0.055)	(0.066)
T4: Female vs. Male	0.063	0.096	0.025	-0.185**
	(0.060)	(0.068)	(0.031)	(0.062)

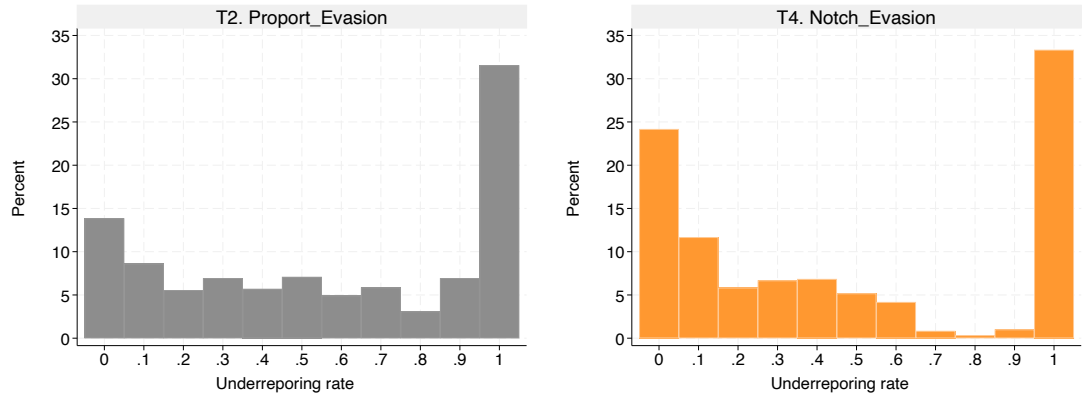
Notes. Standard errors clustered at the subject level are in parentheses. Stars for statistical significance are only reported for marginal effects: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure C.2: Distributions of underreporting rates in treatments 2 and 4, by gender.

(a) - Female



(b) - Male



Appendix D Breakdown of estimates by periods

This appendix presents the results of the statistical analyses conducted in the paper, redone by dividing the 10 payment periods of the experiment into periods 3-9 and the last three periods 10-12. Since participants can learn during experiments, it is relevant to examine whether and, if so, how learning can affect tax notch responses in the present experiment.

Table D.1 reports the multinomial logit estimates for the 5 effort categories, with Tables D.2 and D.3 showing the corresponding predicted probabilities and marginal effects from interactions between treatments and two period intervals. These tables particularly confirm for both period intervals R1: bunching of efforts in treatments 3 and 4 (with more bunching when evasion is not possible), which remain fairly close across the two period intervals in both treatments (see marginal effects in Table D.3). The marginal effects across the period intervals also show the general effect of learning in the slider task, with more effort choices above 18 sliders occurring in all treatments and fewer for effort lower than 10 and between 10 and 15 sliders.

In treatments 3 and 4, the reduced frequencies between 10 and 15 sliders may also derive from a decrease in optimization frictions. In relative terms, the reduced choices are the following: -0.321 (0.065) in treatment 1, -0.279 (0.081) in treatment 2, -0.552 (0.078) in treatment 3, and -0.417 (0.068) in treatment 4. In treatment 3, the greater proportional reduction compared to treatments 1 and 2 ($p<0.05$ in both comparisons) supports a reduction in optimization frictions beyond any learning effect. This is reinforced by the drop in absolute frequency of choices in the region during the last three periods of treatment 3 (at 7%).

In treatment 4, a large proportional decrease in effort is also observed. However, the difference is not significantly different from the other treatments (neither compared to treatments 1 and 2, nor to treatment 3).

In treatment 4, choices between 10 and 15 sliders can be definitively attributed to frictions only when they occur with full compliance. Table D.4 reports the logit estimates for effort at the threshold and between 10 and 15 sliders in treatment 4, with a breakdown between full compliance (FC=1) and evasion (FC=0). Table D.5 shows the corresponding predicted probabilities and marginal effects. The predicted probabilities for effort between 10 and 15 sliders in the last three experimental periods are 20% in case of full compliance and 17% in case of evasion (Tab. D.5, col. 2). The proportions are not negligible, including for full compliance and when compared with those in treatments 1 and 2 (0.23 and 0.22, respectively; D.2, col. 3). Overall, these results confirm across the two period intervals a more robust tendency of frictions in treatment 4, than in

treatment 3 (R2 of the text).

Table D.4 reports the multinomial logit estimates for the eight effort-reported income categories in treatments 2 and 4 and Table D.5 reports the corresponding predicted probabilities and marginal effects. The results show the robustness of the R3 findings across both period intervals. In both intervals, the main tax notch effects in treatment 4 concern the decrease in choices when effort is above 10 and reported income is below 10 (column 6), compensated by the increases in choices of effort at the threshold with full compliance (column 3), and effort above 10 with reported income at 10 (column 7).

Table D.8 shows the multinomial logit model estimates for the underreporting rates in four classes and Table D.9 presents the corresponding predicted probabilities and marginal effects. The results confirm R4 for both time intervals: namely, a statistically significant decrease in large-scale evasion in treatment 4 compared to treatment 2, accompanied by an increase in full-compliance, while the frequency of total evasion choices is unchanged.

Table D.1: Estimated coefficients of effort choices across treatments - Periods 3-9 and 10-12 (Multinomial Logit).

	(1)	(2)	(3)	(4)
	Eff.< 10	Str. domin. Eff. $(10-15]$	Eff. $(15-18]$	Eff. > 18
2. Proport_Evasion	0.013	-0.081	0.258	0.567
	(0.882)	(0.590)	(0.668)	(0.683)
3. Notch_NO_Evasion	-2.636***	-3.329***	-2.459***	-2.165***
	(0.706)	(0.444)	(0.518)	(0.555)
4. Notch_Evasion	-1.298	-1.879***	-1.976***	-1.969***
	(0.690)	(0.490)	(0.562)	(0.597)
Female=1	0.208	-0.396	-0.749	-1.515*
	(0.628)	(0.454)	(0.545)	(0.590)
2. Proport_Evasion × Female=1	0.334	0.087	-0.457	-0.624
	(0.976)	(0.692)	(0.795)	(0.855)
3. Notch_NO_Evasion × Female=1	0.596	0.557	-0.060	-0.097
	(0.815)	(0.539)	(0.629)	(0.701)
4. Notch_Evasion × Female=1	-0.476	0.666	0.766	1.330
	(0.824)	(0.594)	(0.687)	(0.743)
Periods 10-12=1	-0.868*	-0.527	-0.135	0.376
	(0.440)	(0.300)	(0.302)	(0.293)
2. Proport_Evasion × Periods 10-12	1.190	1.124	1.056	1.116
	(0.745)	(0.662)	(0.665)	(0.651)
3. Notch_NO_Evasion × Periods 10-12	0.337	-0.150	0.291	0.353
	(0.497)	(0.368)	(0.346)	(0.322)
4. Notch_Evasion × Periods 10-12	0.053	0.070	0.195	0.225
	(0.578)	(0.377)	(0.367)	(0.345)
Risk	0.017	0.101*	0.100*	0.080
	(0.067)	(0.045)	(0.051)	(0.055)
Constant	1.258	2.596***	2.428***	2.684***
	(0.764)	(0.620)	(0.701)	(0.755)
Controls	Yes	Yes	Yes	Yes
Obs.	5100			
Clust.	510			
Chi-2	587.32			
Prob >Chi-2	0.0000			
Pseudo R-2	0.0794			

Notes. This table reports the coefficients estimated from the Multinomial Logit model for effort choices across treatments, distinguishing between Periods 3-9 and 10-12 (and without a time trend). The predicted probabilities and marginal effects for the two period intervals are in the following Tables D.2 and D.3. The base category for estimation is effort at 10 (bunched). The baseline is T1. Proport_NO_Evasion. Controls are those reported in Table B.2. Standard errors are clustered at the subject level and reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table D.2: Predicted probabilities of effort choices across treatments - Periods 3-9 and 10-12 (from Multinomial Logit model in Table D.1).

	(1) Eff.< 10	(2) Bunching Eff. $= 10$	(3) Str. domin. Eff. $(10-15]$	(4) Eff. $(15-18]$	(5) Eff. > 18
Predicted probabilities					
T1. Proport_NO_Evasion × Periods 3-9	0.08	0.03	0.34	0.30	0.25
	(0.017)	(0.006)	(0.025)	(0.022)	(0.028)
T1. Proport_NO_Evasion × Periods 10-12	0.04	0.03	0.23	0.29	0.40
	(0.015)	(0.010)	(0.028)	(0.027)	(0.035)
T2. Proport_Evasion × Periods 3-9	0.10	0.03	0.30	0.27	0.30
	(0.020)	(0.007)	(0.024)	(0.021)	(0.028)
T2. Proport_Evasion × Periods 10-12	0.06	0.01	0.22	0.25	0.46
	(0.018)	(0.005)	(0.026)	(0.028)	(0.035)
T3. Notch_NO_Evasion × Periods 3-9	0.08	0.24	0.16	0.24	0.28
	(0.016)	(0.027)	(0.016)	(0.018)	(0.026)
T3. Notch_NO_Evasion × Periods 10-12	0.04	0.21	0.07	0.23	0.44
	(0.015)	(0.030)	(0.013)	(0.026)	(0.036)
T4. Notch_Evasion × Periods 3-9	0.07	0.10	0.30	0.25	0.27
	(0.014)	(0.018)	(0.024)	(0.021)	(0.031)
T4. Notch_Evasion × Periods 10-12	0.03	0.10	0.17	0.25	0.46
	(0.009)	(0.021)	(0.024)	(0.027)	(0.036)

Notes. Standard errors clustered at the subject level are in parentheses. (As in the tables in the text, statistical significance is not reported for predicted probabilities).

Table D.3: Marginal effects of effort choices across treatments - Periods 3-9 and 10-12 (from Multinomial Logit model in Table D.1, and predicted probabilities in Table D.2).

	(1) Eff.< 10	(2) Bunching Eff. $= 10$	(3) Str. domin. Eff. $(10-15]$	(4) Eff. $(15-18]$	(5) Eff. > 18
Marginal effects Periods 3-9					
T2 vs. T1	0.020	-0.002	-0.038	-0.030	0.050
	(0.026)	(0.010)	(0.034)	(0.030)	(0.040)
T3 vs. T1	0.001	0.217***	-0.178***	-0.063*	0.023
	(0.024)	(0.027)	(0.030)	(0.029)	(0.038)
T4 vs. T1	-0.015	0.076***	-0.038	-0.044	0.020
	(0.022)	(0.019)	(0.036)	(0.030)	(0.042)
T4 vs. T2	-0.035	0.078***	-0.001	-0.013	-0.029
	(0.025)	(0.019)	(0.034)	(0.029)	(0.042)
T4 vs. T3	-0.016	-0.141***	0.140***	0.019	-0.002
	(0.022)	(0.032)	(0.029)	(0.028)	(0.040)
Marginal effects Periods 10-12					
T2 vs. T1	0.018	-0.022*	-0.013	-0.045	0.061
	(0.023)	(0.011)	(0.038)	(0.039)	(0.049)
T3 vs. T1	0.004	0.183***	-0.158***	-0.068	0.040
	(0.021)	(0.031)	(0.031)	(0.038)	(0.050)
T4 vs. T1	-0.013	0.063**	-0.054	-0.047	0.052
	(0.017)	(0.023)	(0.037)	(0.039)	(0.051)
T4 vs. T2	-0.032	0.085***	-0.041	-0.002	0.010
	(0.021)	(0.022)	(0.035)	(0.039)	(0.050)
T4 vs. T3	-0.017	-0.120***	0.103***	0.022	0.012
	(0.018)	(0.037)	(0.027)	(0.038)	(0.051)
Marginal effects					
P10-12 vs. P3-9					
T1	-0.042**	0.005	-0.108***	-0.004	0.149***
	(0.013)	(0.009)	(0.023)	(0.026)	(0.022)
T2	-0.044***	-0.015	-0.084**	-0.018	0.161***
	(0.012)	(0.009)	(0.027)	(0.029)	(0.022)
T3	-0.039***	-0.030	-0.088***	-0.009	0.166***
	(0.009)	(0.021)	(0.016)	(0.025)	(0.023)
T4	-0.040**	-0.009	-0.125***	-0.007	0.181***
	(0.013)	(0.016)	(0.024)	(0.028)	(0.025)

Notes. Standard errors clustered at the subject level are in parentheses. Statistical significance is indicated as: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table D.4: Estimated coefficients of bunching and strictly dominated effort choices in treatment 4 - Periods 3-9 and 10-12 (Logit models).

	(1) Bunching Eff. =10	(2) Str. domin. Eff. $(10-15]$
FC=1	13.682***	0.823
	(8.250)	(0.282)
Female=1	0.947	1.259
	(0.580)	(0.357)
FC=1 × Female=1	0.528	1.368
	(0.382)	(0.564)
Periods 10-12=1	1.010	0.460***
	(0.414)	(0.081)
FC=1 × Periods 10-12=1	0.864	1.258
	(0.415)	(0.400)
Risk	0.969	1.057
	(0.064)	(0.046)
Controls	Yes	Yes
Obs.	1330	1330
Clust.	133	133
Chi-2	60.98	51.27
Prob >Chi-2	0.0000	0.0000
Log pseudolikelihood	-355.7308	-735.8297

Notes. This table reports the coefficients estimated from the logit models for bunching and strictly dominated effort choices in treatment 4, distinguishing between Periods 3-9 and 10-12 (and without a time trend). The predicted probabilities and marginal effects for the two period intervals are in the following Table D.5. The baseline is FC=0 (choice with evasion), Female=0 ('male' and 'prefer not to say'), and Period 3-9 (no interaction with the gender dummy). Controls are those reported in Table B.2. Standard errors are clustered at the subject level and reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table D.5: Predicted probabilities and marginal effects of bunching and strictly dominated effort choices in treatment 4 - Periods 3-9 and 10-12 (from Logit models in Table D.4).

	(1) Bunching Eff. = 10	(2) Str. domin. Eff. $(10-15]$
Predicted probabilities		
FC=1×Periods 3-9	0.27	0.30
	(0.040)	(0.039)
FC=0×Periods 3-9	0.04	0.30
	(0.010)	(0.027)
FC=1×Periods 10-12	0.24	0.20
	(0.049)	(0.041)
FC=0×Periods 10-12	0.04	0.17
	(0.017)	(0.026)
Marginal effects		
FC=1: Periods 10-12 vs. Periods 3-9	-0.024	-0.097*
	(0.040)	(0.046)
FC=0: Periods 10-12 vs. Periods 3-9	0.001	-0.132***
	(0.015)	(0.028)
Periods 3-9: FC=1 vs. FC=0	0.227***	-0.002
	(0.041)	(0.046)
Periods 10-12: FC=1 vs. FC=0	0.203***	0.033
	(0.051)	(0.046)

Notes. Standard errors clustered at the subject level are in parentheses. Stars for statistical significance are only reported for marginal effects: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table D.6: Estimated coefficients of joint reported income and effort choices in treatments 2 and 4 - Periods 3-9 and 10-12 (Multinomial Logit).

	(1) Eff.<10 FC	(2) Eff.<10 Ev	(3) Eff.=10 Ev	(4) Eff.>10 FC	(5) Eff.>10 Rep.<10	(6) Eff.>10 Rep.=10	(7) Eff.>10 Rep.>10
T4. Notch_Evasion	-1.363	-2.726**	-1.521	-3.149***	-3.408***	-2.011*	-2.993***
	(0.887)	(0.966)	(1.053)	(0.754)	(0.751)	(0.825)	(0.760)
Female=1	1.431	1.047	1.041	-0.414	-0.748	-0.200	-0.489
	(0.837)	(0.916)	(1.114)	(0.827)	(0.843)	(0.920)	(0.872)
T4. Notch_Evasion × Female=1	-1.940	-0.744	-0.813	1.320	0.482	0.726	0.959
	(1.111)	(1.129)	(1.309)	(0.930)	(0.976)	(1.048)	(0.986)
Periods 10-12=1	-0.952	0.685	0.0554	1.020	1.090	0.876	1.011
	(1.428)	(1.135)	(1.276)	(1.090)	(1.105)	(1.131)	(1.101)
T4. Notch_Evasion × Periods 10-12	0.530	-2.043	0.106	-0.875	-0.830	-0.587	-0.907
	(1.504)	(1.310)	(1.352)	(1.111)	(1.127)	(1.158)	(1.122)
Risk	0.0409	0.0862	0.151	0.0109	0.244**	0.112	0.0597
	(0.107)	(0.132)	(0.121)	(0.079)	(0.094)	(0.085)	(0.080)
Constant	0.0766	2.768*	1.977	4.434***	4.483***	3.270*	5.511***
	(1.367)	(1.335)	(1.665)	(1.204)	(1.298)	(1.325)	(1.286)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	2620						
Clust.	262						
Chi-2	423.46						
Prob >Chi-2	0.0000						
Pseudo R-2	0.0879						

Notes. This table reports the coefficients estimated from the Multinomial Logit model for the joint reported income and effort choices in treatments 2 and 4, distinguishing between Periods 3-9 and 10-12 (and without a time trend). The predicted probabilities and marginal effects for the two period intervals are in the following Table D.7. The base category for estimation is effort at 10 with full compliance. The baseline is T2. Proport_Evasion. Controls are those reported in Table B.2. Standard errors are clustered at the subject level and reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table D.7: Predicted probabilities and marginal effects of joint reported income and effort choices in treatments 2 and 4 - Periods 3-9 and 10-12 (from Multinomial Logit in Table D.6).

	(1) Eff.<10 FC	(2) Eff.<10 Ev	(3) Eff.=10 FC	(4) Eff.=10 Ev	(5) Eff.>10 FC	(6) Eff.>10 Rep.<10	(7) Eff.>10 Rep.=10	(8) Eff.>10 Rep.>10
Predicted probabilities								
T2. Proport_Evasion × Periods 3-9	0.04	0.07	0.01	0.02	0.15	0.42	0.05	0.24
	(0.010)	(0.015)	(0.003)	(0.007)	(0.023)	(0.029)	(0.008)	(0.027)
T2. Proport_Evasion × Periods 10-12	0.01	0.05	0.00	0.01	0.16	0.47	0.05	0.25
	(0.006)	(0.017)	(0.003)	(0.004)	(0.028)	(0.035)	(0.014)	(0.031)
T4. Notch_Evasion × Periods 3-9	0.03	0.03	0.08	0.03	0.18	0.26	0.13	0.26
	(0.009)	(0.008)	(0.016)	(0.007)	(0.022)	(0.030)	(0.021)	(0.030)
T4. Notch_Evasion × Periods 10-12	0.02	0.01	0.07	0.03	0.18	0.29	0.16	0.26
	(0.007)	(0.004)	(0.018)	(0.012)	(0.025)	(0.034)	(0.026)	(0.033)
Marginal effects								
T2: Periods 10-12 - Periods 3-9	-0.032***	-0.015	-0.004	-0.011	0.010	0.046	-0.004	0.011
	(0.008)	(0.010)	(0.004)	(0.008)	(0.019)	(0.025)	(0.014)	(0.022)
T4: Periods 10-12 - Periods 3-9	-0.013	-0.025**	-0.009	0.001	0.003	0.031	0.022	-0.008
	(0.009)	(0.009)	(0.012)	(0.011)	(0.018)	(0.021)	(0.019)	(0.024)
Periods 3-9: treatment 2 - treatment 4	-0.006	-0.037*	0.071***	0.009	0.025	-0.166***	0.082***	0.021
	(0.013)	(0.016)	(0.016)	(0.009)	(0.032)	(0.042)	(0.023)	(0.041)
Periods 10-12: treatment 2 - treatment 4	0.013	-0.047**	0.066***	0.021	0.018	-0.181***	0.109***	0.002
	(0.009)	(0.018)	(0.018)	(0.013)	(0.039)	(0.049)	(0.030)	(0.046)

Notes. Standard errors clustered at the subject level are in parentheses. Stars for statistical significance are only reported for marginal effects: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table D.8: Estimated coefficients of underreporing rates in treatments 2 and 4 - Periods 3-9 and 10-12 (Multinomial Logit).

	(1) Full compliance	(2) Medium-large evasion (0.5 <URR< 1)	(3) Total evasion
T4.Notch_Evasion	0.321	-1.106*	0.062
	(0.386)	(0.436)	(0.364)
Female=1	-0.035	-0.192	-0.722
	(0.385)	(0.362)	(0.417)
T4.Notch_Evasion × Female=1	0.032	0.201	-0.399
	(0.487)	(0.576)	(0.580)
Periods 10-12=1	-0.018	0.296*	0.301*
	(0.164)	(0.143)	(0.128)
T4.Notch_Evasion × Periods 10-12=1	-0.005	0.004	-0.136
	(0.206)	(0.261)	(0.188)
Risk	-0.065	0.142*	0.151*
	(0.050)	(0.061)	(0.067)
Constant	-1.180*	-1.174	-1.486
	(0.589)	(0.639)	(0.761)
Controls	Yes	Yes	Yes
Obs.	2620		
Clust.	262		
Chi-2	96.63		
Prob >Chi-2	0.0000		
Pseudo R-2	0.0720		

Notes. This table reports the coefficients estimated from the Multinomial Logit model for the underreporting rates (URR) in treatments 2 and 4, distinguishing between Periods 3-9 and 10-12 (and without a time trend). The predicted probabilities and marginal effects for the two period intervals are in the following Table D.9. The base category for estimation is medium-small evasion (0 <URR≤ 5). The baseline is T2. Proport_Evasion. Controls are those reported in Table B.2. Standard errors are clustered at the subject level and reported in parentheses. Statistical significance is indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table D.9: Predicted probabilities and marginal effects of underreporting rates in treatments 2 and 4 - Periods 3-9 and 10-12 (from Multinomial Logit in Table D.8).

	(1) Full compliance	(2) Medium-small evasion (0 <URR≤ 0.5)	(3) Medium-large evasion (0.5 <URR< 1)	(4) Total evasion
Predicted probabilities				
2. Proport_Evasion × Periods 3-9	0.19***	0.38***	0.20***	0.23***
	(0.025)	(0.030)	(0.027)	(0.029)
2. Proport_Evasion × Periods 10-12	0.17***	0.33***	0.23***	0.27***
	(0.029)	(0.033)	(0.031)	(0.034)
4. Notch_Evasion × Periods 3-9	0.29***	0.40***	0.08***	0.23***
	(0.029)	(0.033)	(0.016)	(0.029)
4. Notch_Evasion × Periods 10-12	0.27***	0.38***	0.10***	0.25***
	(0.032)	(0.036)	(0.022)	(0.033)
Marginal effects				
T2: Periods 10-12 vs. Periods 3-9	-0.024	-0.044	0.032	0.037
	(0.020)	(0.024)	(0.022)	(0.021)
T4: Periods 10-12 vs. Periods 3-9	-0.020	-0.022	0.021	0.021
	(0.022)	(0.024)	(0.018)	(0.021)
Periods 3-9: T4 vs. T2	0.091*	0.026	-0.119***	0.002
	(0.038)	(0.045)	(0.032)	(0.041)
Periods 10-12: T4 vs. T2	0.095*	0.048	-0.130***	-0.014
	(0.044)	(0.049)	(0.039)	(0.048)

Notes. Standard errors clustered at the subject level are in parentheses. Stars for statistical significance are only reported for marginal effects: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.