

Pfeil, Katharina; Kasper, Matthias; Necker, Sarah; Feld, Lars P.

**Working Paper**

## Tax System Design, Tax Reform, and Labor Supply

CESifo Working Paper, No. 11350

**Provided in Cooperation with:**

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Pfeil, Katharina; Kasper, Matthias; Necker, Sarah; Feld, Lars P. (2024) : Tax System Design, Tax Reform, and Labor Supply, CESifo Working Paper, No. 11350, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/305592>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

## Tax System Design, Tax Reform, and Labor Supply

*Katharina Pfeil, Matthias Kasper, Sarah Necker, Lars P. Feld*

## **Impressum:**

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email [office@cesifo.de](mailto:office@cesifo.de)

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: [www.SSRN.com](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: <https://www.cesifo.org/en/wp>

# Tax System Design, Tax Reform, and Labor Supply

## Abstract

This study examines how tax system design and reform affect labor supply. We conduct an online experiment with 522 participants to assess labor responses to tax reforms that introduce or remove a notch, affecting after-tax income at either the lower or upper end of the income distribution. Our findings indicate asymmetric responses to tax reform as well as substantial heterogeneity at the individual level. In particular, we find an increase in labor supply in response to a tax reform only when the reform reduces the tax burden at the upper end of the income distribution. While, in the aggregate, labor supply adjusts on the extensive and intensive margins, we also find strong evidence of heterogeneity in individual responses, showing that the labor response is primarily driven by individuals directly affected by the reform. We examine the role of misperceptions at the individual level as well as fairness considerations in explaining these results.

JEL-Codes: J200, J220, H240, H300, C910.

Keywords: tax system design, tax reform, notches, labor supply, online experiment.

*Katharina Pfeil*  
*Walter Eucken Institute*  
*Freiburg / Germany*  
*pfeil@eucken.de*

*Matthias Kasper*  
*Walter Eucken Institute*  
*Freiburg / Germany*  
*kasper@eucken.de*

*Sarah Necker*  
*Ludwig Erhard ifo Center for Social Market*  
*Economy and Institutional Economics*  
*Fürth / Germany*  
*necker@ifo.de*

*Lars P. Feld\**  
*Walter Eucken Institute & University of*  
*Freiburg / Germany*  
*feld@eucken.de*

\*corresponding author

We appreciate funding from the Wissenschaftliche Gesellschaft Freiburg and thank Zareh Asatryan, Annabelle Doerr, Lilith Burgstaller, Philipp Doerrenberg, Nadja Dwenger, Marcus Giamattei, Martin Fochmann, Friedrich Heinemann, Martin Kocher, Wladislaw Mill, Cornelius Schneider, and Mazhar Waseem, as well as seminar participants from GfW 2022, YEM 2019, WEI's conference on third-party reporting, and seminars in Freiburg for helpful comments. This study was pre-registered at AEA RCT Registry No. AEARCTR-0008410, see <https://doi.org/10.1257/rct.8410> where the data and code is available.

# 1 Introduction

A large body of work studies the effect of taxation on labor supply and the role of behavioral mechanisms in shaping this effect. Prior work finds that the type of tax (Blumkin et al., 2012), tax salience (Hayashi et al., 2013), tax system complexity (Blaufus et al., 2022; Abeler & Jäger, 2015), redistribution (Keser et al., 2020; Soldatos, 2021; He, 2020), tax aversion (Kessler & Norton, 2016), and misperceptions at the individual level (Avram, 2015; Fochmann & Weimann, 2013; Fochmann et al., 2013) affect labor responses to taxation. Less is known, however, about the labor supply effects of progressivity in the income tax systems, as well as tax reforms affecting progressivity. In this paper, we study how tax system design and tax reforms affect labor-leisure tradeoffs at the individual level.

An important feature of many progressive tax systems are discontinuous changes in the tax burden, or notches, at specific income thresholds. These discontinuities often result from tax policies targeted at low-income earners such as the Earned Income Tax Credit in the United States (Bastian & Lochner, 2022), the Minijob scheme in Germany (Tazhitdinova, 2020), the marginal employment threshold in Austria (Steiner & Wakolbinger, 2013), labor vouchers in Italy (Di Porto et al., 2022), the Working Family Tax Credit and zero-hours contracts in the UK (Dolado et al., 2022; Azmat, 2014), and casual work in Australia (Scarfe, 2019). Thresholds are also used to introduce progressivity in other settings, such as personal income taxes (Kleven & Waseem, 2013), corporate income taxes (Bachas & Soto, 2018), or value added taxes (Liu & Lockwood, 2015; Liu et al., 2021).<sup>1</sup>

Since notches in the tax system induce a region of strictly dominated choice, individuals can increase their consumption and leisure by bunching below the threshold (Kleven & Waseem, 2013).<sup>2</sup> However, prior work shows that optimization errors or frictions can undermine people’s ability to respond optimally to discontinuities in the tax system (see, e.g., Kleven, 2016; Kostøl & Myhre, 2021). Moreover, even in the absence of frictions, many individuals do not bunch at optimal levels, suggesting that behavioral factors, such as self-control, loss aversion, and overconfidence affect individual decision making (Gibson et al., 2019). More broadly, individual responses to changes in the tax system may also deviate from standard theoretical predictions,

---

<sup>1</sup>See Kleven (2016), Slemrod (2013), Saving & Viard (2021) for detailed overviews of notches in tax systems.

<sup>2</sup>Economists have long debated the costs and benefits of notches and pointed to (labor) distortions (Blinder & Rosen, 1985). The political intention of notches is to motivate particular groups to change their behavior (Slemrod, 2013). Governments also use thresholds to screen out individuals with higher work capacity or better health (Krekó et al., 2022).

because individuals are inattentive or lack knowledge (Amberger et al., 2023; Chetty et al., 2013; Chetty & Saez, 2013; Chetty, 2012), or because search and renegotiation costs and adjustment frictions may attenuate individuals’ ability to adapt to changes in tax incentives (Tazhitdinova, 2020). In line with these arguments, Kostøl & Myhre (2021) attribute two-thirds of such frictions to misperceptions of the tax code.

A related strand of literature investigates the effect of tax reforms on labor supply and several studies document behavioral responses that deviate from the predictions of standard economic theory. For example, Doerrenberg & Duncan (2014) examine the effect of tax-rate changes in flat tax systems and find that the evolution of such tax changes affects labor provision. Another study by Pántya et al. (2016) finds that work performance increases significantly when a flat tax system is transformed into a progressive system. Similarly, Masclet & Montmarquette (2008) find that a progressive tax system leads to less labor supply than a proportional system and that individuals’ previous tax experiences affect their labor supply decisions. With regard to reforming discontinuities in the tax system, prior work finds that behavioral responses lag reforms so that bunching at new thresholds builds over time, while bunching at old thresholds diminishes gradually, rather than instantly (Mavrokonstantis & Seibold, 2022).

Despite this large body of work on the effect of taxes on labor supply, several important issues remain unclear. In particular, it remains unclear how labor supply responds to the introduction or elimination of progressivity in taxation and to what extent the effects of tax reform hinge on the design of the pre-reform tax system. Moreover, it also remains unclear to what extent aggregate responses to tax reform are driven by individuals who are directly affected by the reform. Similarly, prior work largely ignores individuals who are unaffected by the reform, but might still adjust their labor supply due to the biases and adjustment frictions identified in prior work.

Our study uses an online experiment to address these important gaps in the literature. In particular, we investigate the labor supply effects of tax system design and tax reform. We study whether labor responses depend on the type of reform, i.e., a tax cut or a tax increase that results from the transformation of a flat tax system into a progressive tax system via the introduction, or elimination, of a notch. Moreover, we investigate path dependency in labor responses to tax reform by analyzing how the pre-reform tax system affects post-reform labor responses. Finally, we explore to what extent aggregate responses mask heterogeneity among individu-

als who are affected versus unaffected by the reform. Addressing these issues in a controlled online setting allows us to hold constant other policy changes that may accompany a tax reform and to rule out confounding effects of frictions, such as lack of information, workers' inability to negotiate hours, or career concerns. Finally, it allows us to investigate in detail the effects of tax reform resulting in tax increases or tax cuts on the overall sample as well as sub populations of (un-)affected individuals.

In our study, we implement a real-effort experiment in an online labor market (Amazon's Mechanical Turk), where participants are randomly assigned to either one of four treatment conditions or one control condition. Participants then face 16 labor supply decisions in each condition. We introduce variation in two dimensions. First, we vary the type of tax reform that is introduced after the first half of the experiment. Specifically, in some treatments a flat tax system is transformed into a progressive tax system via the introduction of a notch. In another treatment, a progressive system is transformed into a flat system via the elimination of a notch. Second, we experimentally manipulate which part of the income distribution is affected by the reform by varying whether the reform implements a tax cut or a tax increase above or below the threshold. We also include a control treatment with a progressive tax system and no tax reform. Our predictions are threefold. First, we expect that tax reforms affect labor supply. Second, we anticipate that labor responses are either driven by an income effect, respectively loss aversion, or a substitution effect and that the pre-reform tax system affects post-reform labor responses. Third, we expect that labor supply responses depend on individual affectedness and preferences for income over leisure.

Our findings suggest that tax reforms that decrease the tax burden lead to an increase in labor supply, suggesting that a substitution effect, rather than an income effect, drives behavioral responses. In summary, however, our results reveal nuanced labor responses to tax reform. Specifically, labor provision is affected only by a reform that decreases the tax burden by transforming a progressive into a flat tax regime. In this case, subjects prefer maximizing income over leisure in the post-reform periods. In contrast, we do not find evidence of labor supply effects of tax reforms increasing the tax burden nor do we find effects of reforms resulting in a decrease of the tax burden due to the transformation of a flat tax into a progressive tax regime. These findings indicate path dependency in behavioral responses to tax reform. In particular, we find that post-reform labor responses hinge on the design of the pre-reform tax system.

A more detailed analysis of these results reveals that the absence of labor supply effects in the aggregate masks heterogeneous responses of individuals who increase, or decrease their labor supply after experiencing a tax reform and that behavioral responses are driven mostly by individuals who are affected by the reform. Moreover, we find strong labor supply effects of tax reforms on the extensive and intensive margins when a tax reform replaces a progressive tax system with a flat tax system resulting in a decrease in the tax burden. Finally, a post-experimental survey reveals that subjects are more likely to optimize labor supply, consider the reform fairer, and compute the tax burden more accurately when the tax reform decreases the tax burden by transforming a progressive into a flat tax regime.

Our study contributes to the literature on several dimensions. We extend prior work finding asymmetric effects of changes in wages on labor supply. Unlike Doerrenberg et al. (2023), who show that wage increases have smaller effects on labor supply than wage decreases of equal magnitude, our findings indicate that different responses to increases and decreases in wages cannot convincingly be explained by loss aversion. Moreover, our design allows us to rule out learning effects or fatigue as potential explanations of asymmetry in behavioral responses to tax reform (Doerrenberg & Duncan, 2014). Instead, our findings suggest that tax reforms affect labor responses only when they reduce the tax burden at the upper end of the income distribution, i.e., above the notch inducing progressivity in the income tax system. We also extend prior work showing that individuals' prior tax experiences affect labor supply decisions (Masclot & Montmarquette, 2008) and provide a novel perspective on the drivers of stickiness in behavioral responses to changes in thresholds in the tax system (Mavrokonstantis & Seibold, 2022).

Our findings inform the debate on implications of fiscal stimulus programs aimed at increasing investment or consumption using temporary subsidies or tax cuts (e.g., Agarwal et al., 2007; Mian & Sufi, 2012; Johnson et al., 2006). Taken together, our results suggest that behavioral responses to tax reform do not necessarily align with theoretical predictions and that individuals do not optimally react to the reform of thresholds in the tax system even in a world without frictions. In order to achieve the intended effects of a tax reform, policymakers should thus not only focus on the design of the reform, but also on behavioral aspects, such as path dependency and fairness concerns.

This paper is structured as follows. In Section 2, we describe the experimental design. Section 3 develops theoretical expectations. Section 4 presents the results.



Section 5 concludes.

## 2 Experimental design

### 2.1 Experimental procedure

We examine the effect of tax reforms on labor supply in an online experiment. This setting allows us to introduce systematic variation in the type and timing of tax reforms, as well as in the amount of income that participants can earn by providing labor. Moreover, an online experiment provides a suitable environment to model labor-leisure decisions as participants can independently decide on their leisure once they have completed the study.<sup>3</sup> The study was implemented using Amazon Mechanical Turk (MTurk), an online crowdsourcing platform that connects employers with workers to perform tasks. Human Intelligence Tasks (HIT) are posted on the platform for potential workers to search and complete. Workers can complete any HIT for the amount of monetary compensation offered for the task.

The flow of our experiment is shown in Figure 1. Appendix B includes a detailed overview of the experimental design. When participants enter the study, they first have to fill in a brief survey on sociodemographic characteristics. Participants then proceed to the instructions and answer a set of control questions. Those who do not answer correctly within two attempts are redirected to the instructions. Subsequently, participants proceed to the main experiment which comprises 16 periods. In each period, they decide how much labor to provide given the tax rate that applies in the respective period (“decision stage”). Once participants have completed all 16 periods, one period is randomly selected and participants have to complete the number of tasks they indicated in that period (“working stage”).<sup>4</sup> Finally, subjects fill in a post-experimental survey which inquires information on preferences and perceptions. At the end of the experiment, participants receive a show-up fee and an additional payment based on their performance in the randomly selected period. The implementation of the experiment is detailed in Section 2.4 below.

---

<sup>3</sup>There have been various implementations of leisure in laboratory experiments (e.g., Abeler et al., 2011; Corngnet et al., 2015; Eriksson et al., 2009; Mohnen et al., 2008; Dickinson, 1999).

<sup>4</sup>Two participants exit the experiment on the working stage.

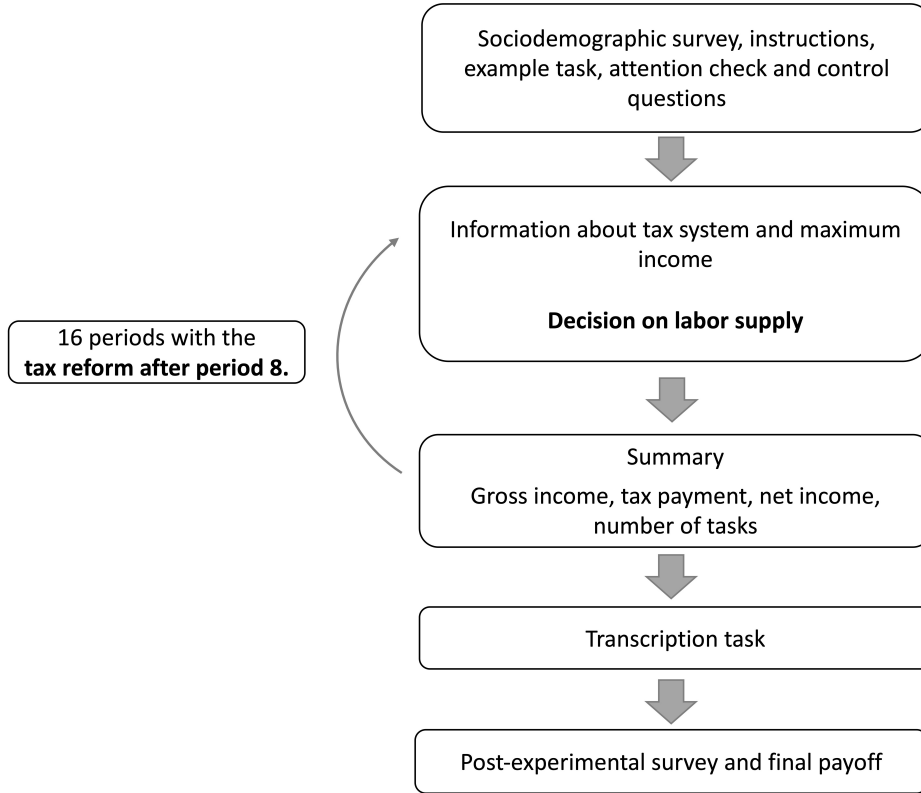


Figure 1: Experimental procedure

## 2.2 Decision stage: Treatments

Participants are randomly assigned to one of five experimental treatments that differ only with regard to the decision stage. Each period of the decision stage is structured as follows. On the first screen, we display how income is taxed (between-subject variation) as well as the maximum income participants can earn by providing labor (i.e., by solving tasks) in the respective period (within-subject variation). On the same screen, participants decide on the income they want to earn and the corresponding amount of labor that they need to provide. The maximum gross income is 600 cents and each task pays 20 cents so that participants can provide up to 30 labor units by solving up to 30 tasks. For each gross income, the corresponding amount of tasks as well as the resulting tax payment are automatically calculated and displayed. The lower the income that participants choose, the fewer tasks they need to solve. Thus, participants face a trade-off between earning more money and enjoying more leisure. Once participants have made their decision, they proceed to the summary screen

which displays the chosen gross income, the tax payment, the number of tasks required to earn the income and the potential payoff if this period is randomly chosen.

### 2.2.1 Treatment design

The five experimental treatments are presented in Table 1. In the baseline treatment *Prog, Prog*, subjects make decisions in a progressive tax system for 16 periods. This treatment does not include a tax reform. The progressive tax system is designed as follows. Gross incomes of up to 400 cents (20 labor units) are taxed at a tax rate of 25%. However, the entire income is taxed at 50%, if the income exceeds the threshold. The baseline treatment thus implements a progressive tax system with a notch at 400 cents. In particular, the instructions state: “In this round, the tax rate is 25% for incomes equal to or below 400 cents. The tax rate is 50% on the entire income if the income exceeds 400 cents. For example, for an income of 420 cents, your tax payment would be 210 cents.”

In the other four treatments, we implement a tax reform that occurs after eight of the overall sixteen periods. In *Prog, Flat25*, we introduce a tax reform that reduces the tax rate. Specifically, participants in this treatment face the progressive tax system described above for eight periods. After the tax reform, i.e., from period nine onwards, all incomes are taxed at a flat rate of 25%. Therefore, the tax reform reduces taxes on income above the gross income threshold of 400 cents (or 20 labor units) by 25 percentage points, while income below that threshold is unaffected by the reform. The respective instructions read: “In this round, the tax rate is 25% for all incomes.” Therefore, the tax reform in this treatment reduces the tax burden considerably.

In *Prog, Flat50*, participants face a tax reform that increases taxes to a flat rate of 50% on all incomes. Consequently, all income below the threshold of 400 cents are subject to a tax increase, while all income above that threshold is unaffected by the reform. The tax reform is announced as follows: “In this round, the tax rate is 50% for all incomes.” The reform in this treatment therefore increases the tax burden.

The remaining treatments implement the same tax reforms in reverse order. Specifically, in *Flat25, Prog*, participants first face a flat tax system with a tax rate of 25%, which is reformed into a progressive tax system after eight periods. In this treatment incomes above the gross income threshold of 400 cents, or 20 tasks, experience a tax increase of 25 percentage points. Finally, in *Flat50, Prog*, subjects start

with eight periods of a flat tax rate of 50%, after which the progressive tax system is introduced. This reform implies a tax decrease for incomes below the threshold in the post-reform period.

No reform	<i>Prog, Prog</i>	
	N = 102	
Tax decrease	<i>Prog, Flat25</i>	<i>Flat50, Prog</i>
	N = 107	N = 99
Tax increase	<i>Prog, Flat50</i>	<i>Flat25, Prog</i>
	N = 104	N = 110

Table 1: Treatment overview

Notes: For each treatment, the number of participants is displayed. Each participant makes 16 decisions. The tax reform occurs after eight of sixteen periods.

### 2.2.2 Variation in income

To increase participants' involvement in the experimental task, we randomly assign one of eight maximum gross incomes in each period. Gross incomes vary between 280 cents and 600 cents (see Table 2 for details). Participants decide in each period how much income to earn by providing the respective amount of labor. The gross income translates directly into the number of text sequences a subject needs to transcribe in the real effort task. 20 cents are equal to one task, so the minimum number of possible tasks is 0, and the maximum number of tasks is 30 (for a gross income of 600).

		<i>Prog</i>		<i>Flat25</i>		<i>Flat50</i>	
$Tasks_{max}$	$Y_{max}$	$Y_{net}$	$t$	$Y_{net}$	$t$	$Y_{net}$	$t$
14	280	210	0.25	210	0.25	140	0.5
16	320	240	0.25	240	0.25	160	0.5
20	400	300	0.25	300	0.25	200	0.5
21	420	210	0.5	315	0.25	210	0.5
22	440	220	0.5	330	0.25	220	0.5
24	480	240	0.5	360	0.25	240	0.5
25	500	250	0.5	375	0.25	250	0.5
30	600	300	0.5	450	0.25	300	0.5

Table 2: Maximum tasks, gross and net incomes, and tax rates by tax regime

Notes: We randomly assign one of the maximum amount of tasks/ income to each period. Subjects can choose any amount of tasks between 0 and  $Tasks_{max}$  in each period.  $Y_{max}$  is the maximum before-tax income in cents that a subject can choose in each period.  $Y_{net}$  denotes the respective after-tax income in cents for each tax treatment depending on the tax rate.

## 2.3 Working stage: Task

To earn income, participants provide labor units by solving real-effort tasks. The task is to transcribe text sequences of ten randomly selected letters, which is similar to the tasks used by Augenblick et al. (2015) and Dickinson (1999). Subjects can complete this simple typing task without previous knowledge or mathematical skills.

The number of tasks that subjects have to complete depends on the choices in the decision stage. Specifically, after completing the decision stage, one of the 16 periods is randomly selected, and subjects have to complete the number of tasks chosen in that period. A task is complete when participants correctly type in every letter. When participants mistype, a new sequence appears. There is no time limit to complete the required number of tasks. We block the copy-paste function to avoid that subjects finish the task without effort. During the instructions, subjects work on an example task to familiarize themselves with this tasks.

## 2.4 Implementation

The experiment was programmed using *LIONESS Lab* (Giamattei et al., 2020). We ran the experiment on MTurk between September 12 and 14, 2022.

We recruited the highest-ranked workers via CloudResearch to ensure high-quality data (Litman et al., 2017). Implementing an experiment online can lead to a trade-off between the lower level of control and a more diverse and experienced subject pool with a real labor-leisure trade-off. In line with the literature, we restrict our sample to participants who are US citizens, with a geographical location in the US, and only allow workers with an approval rate of 95% and more than 500 completed HITs (see Arechar et al., 2018; Peer et al., 2014). We choose US citizens because they comprise the largest share of workers on MTurk (Difallah et al., 2018).

MTurk is sometimes criticized for the participation of bots or less attentive participants who may bias the data (e.g., Cuskley & Sulik, 2022; Webb & Tangney, 2022; Hauser et al., 2022). However, Prissé & Jorrat (2022) find that for several standard economic tasks, online experiments result in identical findings as standard lab experiments and are thus not limited in their validity. Nonetheless, we take several precautions to ensure the highest quality of data possible. First, we control for bots with a captcha test that stops non-human subjects from entering the experiment. Second, we include an attention check to filter out inattentive subjects (Schneider & Mill, 2022). Third, we implement a set of pre-experimental control questions to screen subjects for their understanding of the instructions. Subjects who do not complete the control questions in two attempts are excluded from the experiment. Fourth, we prevent duplicate participation. Finally, during the implementation of the experiment, we closely monitored participants’ feedback as well as worker reviewer platforms such as Turkerview. We did not receive negative feedback on the experimental design or detect any discussions about our experiment.

Participants receive a fixed participation fee of 1 USD if they complete the study. In addition to that, they earn a flexible bonus payment depending on their labor provision in a randomly selected period. The bonus is equivalent to the after-tax income in that period and between 0 USD and 4.50 USD, so that the maximum compensation is 5.50 USD. Participants were paid for their participation within three days after completing the experiment.

### 3 Expectations

Standard economic theory predicts that individuals respond to financial incentives. Labor supply should thus reflect income tax rates and individuals provide the amount of labor that maximizes their after-tax income. Table 3 shows the amount of labor

provision (solved tasks) that maximizes after-tax income.

<i>Tax system</i>	<i>Tax rate if <math>\leq 20</math> tasks</i>	<i>Tax rate if <math>&gt; 20</math> tasks</i>	<i>Number of tasks to maximize income</i>
<i>Prog</i>	25%	50%	20; 30
<i>Flat25</i>	25%	25%	30
<i>Flat50</i>	50%	50%	30

Table 3: Tax systems and income maximizing behavior

Notes: In *Prog*, a tax rate of 25% applies if the participant solves 20 tasks or less. If the participants solves more than 20 tasks, a tax rate of 50% applies to all income.

However, as incentives change, so should labor supply. We therefore predict that:

*H1*: Changes in the income tax rate affect labor provision.

How an increase in income taxes affects labor supply is theoretically ambiguous, however. Specifically, as increasing taxes reduces after-tax income, taxpayers might increase their labor provision to maintain their disposable income (income effect). An alternative explanation for this result is loss aversion (Kahneman & Tversky, 1979). Since individuals perceive a loss in after-tax income as painful, they increase their labor supply when income taxes rise. Conversely, because increasing income taxes makes leisure relatively more attractive than labor, taxpayers might reduce their labor provision and consume more leisure (substitution effect). Our study design allows us to investigate the role of income and substitution effects in labor supply responses to income tax reform. Specifically, we investigate the following competing hypotheses:

*H2a*: Increasing (reducing) the income tax rate has a positive (negative) effect on labor provision (income effect/loss aversion)

*H2b*: Increasing (reducing) the income tax rate has a negative (positive) effect on labor provision (substitution effect)

However, aggregate responses to changes in the income tax rate might conceal more nuanced effects of policy reform. In particular, we assume that the behavioral re-

sponse to a policy reform depends on (i) the nature of the reform – i.e., an increase or a decrease in the tax rate, (ii) individual affectedness – i.e., whether or not the tax reform affects an individual given pre-reform labor supply, and (iii) individual preferences for leisure versus income. For example, an increase in the income tax rate affects the entire population in a flat tax system. In contrast, an increase in the marginal tax rate for high income earners affects the upper, but not the lower end of the income distribution. Moreover, we assume that individuals who are affected by a change in tax rates adjust their labor supply in line with their preferences for income versus leisure, while unaffected individuals do not change their labor supply. Specifically, we predict that individuals who prefer income over leisure (*I-Types*) aim to increase their *T1* income (specific income effect). Conversely, we predict that individuals who prefer leisure over income (*L-Types*) aim to maintain their pre-reform (*T0*) income in the post-reform (*T1*) periods (specific substitution effect). These considerations motivate our final hypothesis:

*H2c*: Behavioral responses to tax policy reform depend on the nature of the reform, individual affectedness, and individual preferences for leisure versus income.

Table 4 summarizes our predictions for *H2c*. In each treatment, we distinguish between individuals who provide  $\leq 20$  and  $> 20$  labor units in *T0*, because the tax rate increases at this threshold in the *Prog* treatment. We expect that individuals who are unaffected by tax policy reform, i.e., individuals who do not experience a change in tax rates given their pre-reform labor supply, do not alter their post-reform labor supply. The only exception applies to *I-Type* individuals in the *Prog,Flat25* treatment, where the provision of both, 20 and 30 labor units maximizes after-tax income in *T0*. As the decision to provide 20 labor units in *T0* could reflect intent to maximize income as well as preferences for providing a moderate amount of labor, we cannot distinguish between *I-Types* and *L-Types* in this case *T0*. However, even when some *I-Type* individuals decide to provide 30 labor units in *T0*, we anticipate that the average labor supply response of *I-Type* individuals in *T1* is positive.

However, we expect that individuals who are affected by tax policy reform, i.e., individuals who experience a change in tax rates given their *T0* labor supply, alter their *T1* labor supply. The specific response to an increase in the tax rate hinges on the nature of the reform. In particular, in *Flat25,Prog*, where taxes increase for individuals who provide  $> 20$  labor units in *T0*, pre-reform income levels are



unattainable in  $T1$ . Therefore, we anticipate that labor supply declines in  $T1$ , independent of preferences for income versus leisure (substitution effect). Conversely, in *Prog,Flat50*, taxes increase for individuals who provide  $\leq 20$  labor units in  $T0$ . Here, individuals who aim to increase their income (*I-Types*), as well as individuals who aim to maintain their pre-reform income (*L-Types*) increase their labor supply in  $T1$  (income effect).

When taxes decrease, we also expect that behavioral responses depend on preferences for income or labor. More specifically, we anticipate that *I-Types* increase their labor supply when taxes decrease (substitution effect), while (*L-Types* reduce their labor provision in this case (income effect). This applies to individuals who provide in  $T0 > 20$  labor units in *Prog,Flat25*, as well as to individuals who provide in  $T0 \leq 20$  labor units in *Flat50,Prog*.

<i>Treatment</i>	<i>Mean labor supply in T0 (units)</i>	<i>Change in tax rate (T0–T1)</i>	<i>Predicted labor supply in T1</i>	
			<i>I-Type</i>	<i>L-Type</i>
<i>Prog,Prog</i>	$\leq 20$	none		no change
	$> 20$	none		no change
<i>Prog,Flat25</i>	$\leq 20$	none	increase	no change
	$> 20$	decrease	increase	decrease
<i>Prog,Flat50</i>	$\leq 20$	increase		increase
	$> 20$	none		no change
<i>Flat25,Prog</i>	$\leq 20$	none		no change
	$> 20$	increase		decrease
<i>Flat50,Prog</i>	$\leq 20$	decrease	increase	decrease
	$> 20$	none		no change

Table 4: Expected labor supply responses by reform affectedness and preferences

Notes: The tax rate increases at 20 labor units in the *Prog* tax system. We define individuals who prefer income over leisure as I-Types and individuals who prefer leisure over income as L-Types. *I-Types* aim to increase their T0 income in T1, *L-Types* aim to maintain their T0 income in T1.

## 4 Results

### 4.1 Sample and data quality

In total, 879 US citizens with an approval rate of  $\geq 95\%$  clicked on the link to enter our study. The final sample comprises data from 522 subjects who make 16 decisions each. Therefore, the following analyses are based on a sample of 8,352 observations.<sup>5</sup> An attrition analysis shows that there are no significant correlations between sociodemographic characteristics and the probability of dropping out of the experiment (Table 8 in Appendix A). Notably, the number of dropouts does not vary significantly between treatments.

On average, subjects complete the experiment in 20 minutes. The average earning is 3.47 USD, equivalent to an hourly wage of 10.65 USD. The mean age in our sample is 42 years, and 58% of the subjects are female. 60% of our sample have at least a bachelor’s degree, 59% report an income below the US median household income in 2022; 49% are full-time employed. Table 9 in Appendix A displays the distribution of sociodemographic characteristics by treatment. By and large, there are no significant differences in sociodemographic characteristics between treatments.<sup>6</sup>

We implement several tests to check the data quality, i.e., whether subjects understand the experimental instructions and rationally maximize their income according to the incentive structure of the experiment. An important measure is the fraction of subjects that choose to provide 20 labor units or less, as subjects can maximize labor and leisure by choosing 20 labor units in the progressive tax regime (see Section 3). We find that 80% of subjects in our baseline setting, the *Prog, Prog* treatment, choose to provide 20 labor units or less. As expected, this fraction is considerably smaller in the tax reform treatments, where it revolves around 66%. These results provide initial indication that labor provision decisions align with the incentive structure of the experiment.

To further investigate this issue, we compare pre- and post-reform behavior in Table 5. We find no significant difference in labor supply decisions below or equal to 20 labor units in *Prog, Prog*, confirming that, absent a policy reform, participants do

---

<sup>5</sup>344 subjects were excluded or dropped out before entering the main experiment (39.1%, Figure 8 in Appendix A). We excluded 16 subjects who did not complete the captcha test, 104 inattentive subjects, and 137 subjects with insufficient understanding who failed the control questions twice. Moreover, 30 subjects left during the instructions, and 57 dropped out on their own terms at the exemplary real-effort task. In addition, 13 subjects terminated the experiment while making decisions or working on the real-effort task.

<sup>6</sup>The only exception is the fraction of respondents reporting a below-median household income, which varies significantly between *Prog, Flat25* and *Flat25, Prog*. We control for sociodemographic characteristics in the regressions.

not alter their labor provision in the course of the experiment. In other words, the vast majority of subjects behaves as a rational agent should. Importantly, however, in the treatments with a tax reform, we find substantial differences in labor supply before and after the reform. This indicates that subjects adapt their behavior in line with changes in incentives: Fractions of labor decisions equal to or below 20 units range between 79% and 88% in progressive tax regimes, and are considerably lower in flat tax regimes, where the provision of more labor translates into higher after-tax earnings. In summary, our initial analysis strengthens our confidence in the quality of our data.

Treatment	Pre-reform	Post-reform	<i>p-value</i> (MWU)
<i>Prog,Prog</i>	.78	.81	0.1084
<i>Prog,Flat25</i>	.88	.46	< 0.0001
<i>Prog,Flat50</i>	.83	.54	< 0.0001
<i>Flat25,Prog</i>	.53	.79	< 0.0001
<i>Flat50,Prog</i>	.54	.79	< 0.0001

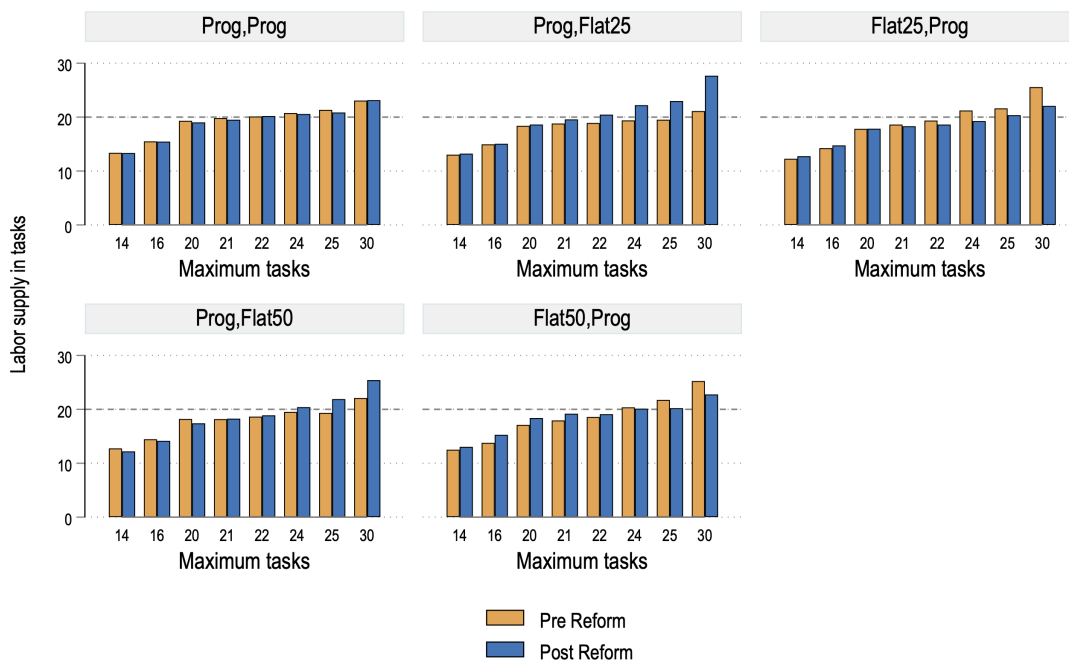
Table 5: Fraction of labor supply decisions below or equal to 20 labor units by treatment

Notes: Pre-reform are periods 1 to 8, and post-reform are periods 9 to 16. The p-values are derived from Mann-Whitney U tests between pre- and post-reform labor supply.

## 4.2 Descriptive analysis

To gain a first intuition as to whether subjects decide according to our expectations detailed in Section 3, we examine labor provision by treatment. Specifically, we investigate whether participants optimize their after-tax income by providing 20 labor units in the progressive tax regime, and maximize their after-tax income by choosing 30 labor units in flat tax regimes. We display the average number of labor units provided over the maximum number of possible labor units in Figure 2. As expected, we find that the average labor supply is at most 20 tasks in the progressive tax regime. In line with financial incentives, we also find the largest share of decisions to provide 30 labor units in *Prog,Flat25*, with approximately 28 tasks. In all other treatments with flat tax regimes, this ratio is around 25. These results indicate that

participants respond as expected to changes in incentives. However, they do not fully maximize their after-tax earnings. Specifically, we find across all treatments that when facing the opportunity to provide 30 labor units, subjects hardly choose to maximize their labor supply.



Graphs by Treatment

Figure 2: Average labor supply (chosen labor units) by maximum possible units

Notes: This figure shows the average number of chosen labor units in tasks for each assigned maximum possible number of tasks. Periods 1 to 8 are pre-reform, and periods 9 to 16 are post-reform.

To investigate in more detail the effect of tax reform, we define our primary outcome variable as the number of tasks provided over the maximum number of tasks possible to provide (in percent). Specifically, in each of the 16 experimental decisions (henceforth: periods), subjects can choose to solve a number of tasks to generate income. Each task yields a gross (pre-tax) income of 20 cents. However, in every period there is a maximum number of tasks that the participant may solve. For that reason, it is not possible for participants to obtain the highest possible income of 600 cents (by solving 30 tasks) in each experimental period. Therefore, we define labor supply as the ratio of labor units (tasks) provided over the maximum number of labor units possible to provide. By normalizing the variable in this way,

we increase comparability of the decisions between periods with varying maximum tasks and are able to observe relative changes in labor.

On average, subjects provide 87% of the maximum possible labor in our sample. In 60% of all decisions, subjects provide the maximum amount of labor possible, resulting in the highest possible gross income. While the mean level of labor supply is highest in *Prog,Prog*, with 90%, it is lowest in *Prog,Flat50*, with 86%.

Distributions of labor provision decisions are highly skewed to the left and differ by treatment, as shown in the density plots depicted in Figure 3. In *Prog,Prog*, relatively few subjects maximize their labor supply and the distribution of labor provision is more dispersed in the upper half. As expected, there is no difference in labor provision between the first (pre-reform) and second (post-reform) half in our baseline treatment. In *Prog,Flat25*, pre- and post-reform labor supply differs. Specifically, the distribution is spread between 60% and 100% of labor supply in pre-reform periods. Post-reform, in the flat tax regime of 25%, labor supply is more concentrated at 90% or up. In the other reform treatments, we see strong bunching of labor supply around 100% in the flat tax regimes, whereas labor supply is more dispersed in the upper half of the distribution in the periods with progressive tax regimes.

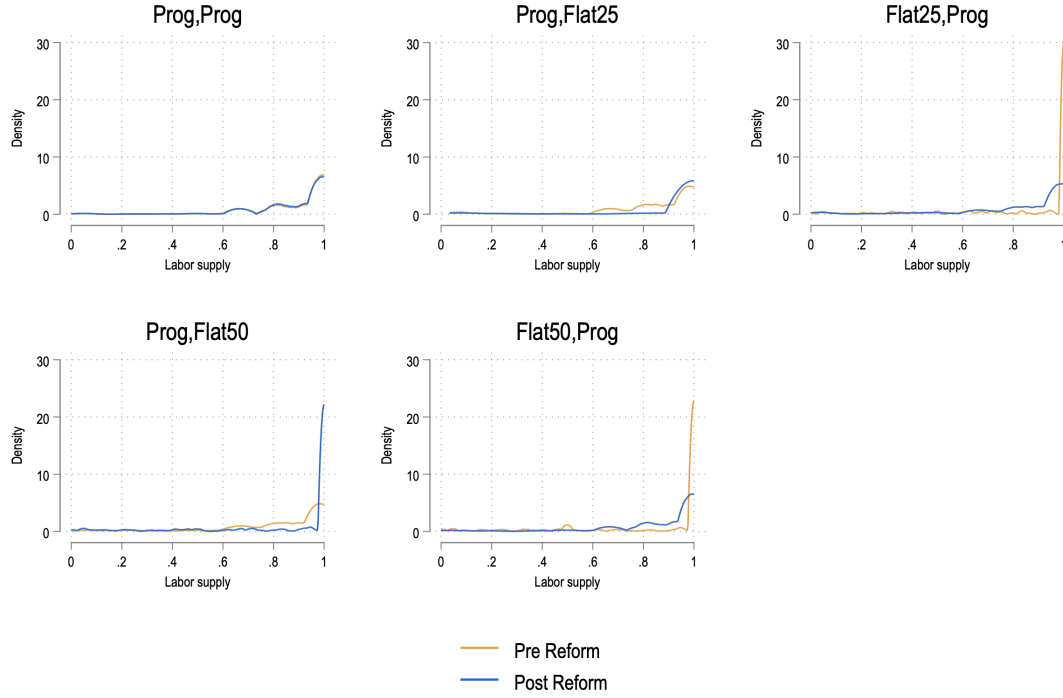


Figure 3: Distribution of labor supply decisions by treatment

Notes: This figure shows kernel density plots of labor supply (in %) by treatment. Periods 1 to 8 are pre-reform, and periods 9 to 16 are post-reform.

For a descriptive intuition of reform effects, we analyze how labor supply evolves over time. Figure 4 shows labor supply by period and treatment. It is evident that labor supply is constantly high, at a level of 80% to 95%. Subjects in our baseline treatment (*Prog,Prog*) do not change their labor provision notably over time. In contrast, labor supply increases after the tax reform introducing a flat tax with a low tax rate in *Prog,Flat25*. Conversely, in *Flat25,Prog*, we observe the reverse, a decline in labor supply after the reform into the progressive regime. Finally, in *Prog,Flat50* supplied labor tends to remain on the same level with no visible reform effect, which is similar in *Flat50,Prog*.

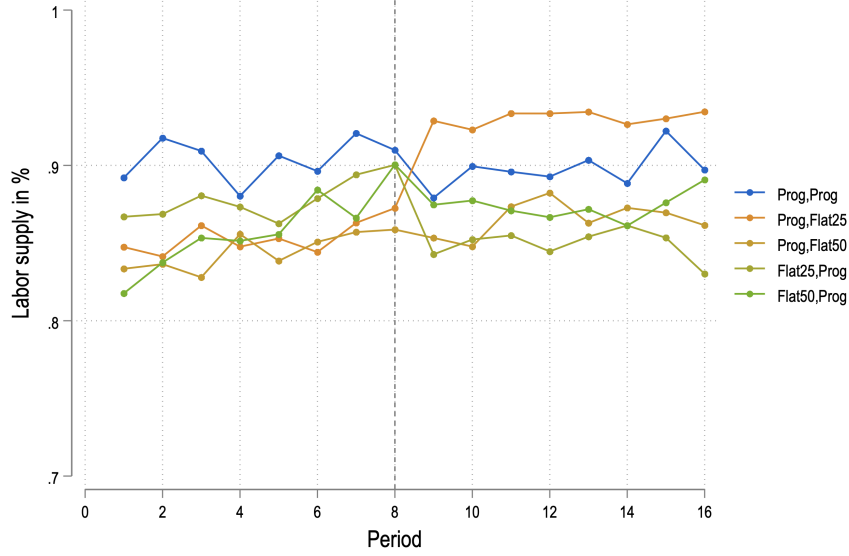


Figure 4: Development of labor supply by period and treatment

Notes: This figure shows mean labor supply (in %) by treatment over 16 decision periods. The tax reform occurs after period eight (dashed line).

### 4.3 Main results

In Section 3, we have hypothesized that changes in income tax rates affect labor supply ( $H1$ ). Increasing (reducing) the income tax rate should affect labor supply positively (negatively) if an income effect or loss aversion determine the labor provision response to tax reform ( $H2a$ ). In contrast, increasing (reducing) taxes should affect labor negatively (positively) if a substitution effect dominates ( $H2b$ ). The descriptive analyses presented above suggest that tax reforms affect labor supply. In particular, we find indication of a substitution effect as predicted in  $H2b$ . When taxes increase, leisure becomes relatively more attractive so that subjects work less. In turn, when taxes decrease, leisure is less attractive and subjects work more. The substitution effect is strong in *Prog,Flat25*, where the reform decreases the income tax rate. Similarly, the tax increase in *Flat25,Prog* reduces labor supply. We next investigate whether these results hold when we compare pre- to post-reform labor supply relative to the baseline treatment, accounting for sociodemographic characteristics.

To test the causal effect of tax policy reform on labor supply more formally, we

estimate the following difference-in-differences model:

$$Labor\ Supply_{it} = \beta_0 + \beta_1 \cdot Post_t + \beta_2 \cdot Treat_i + \beta_3 \cdot (Post_t \times Treat_i) + \epsilon_{it}$$

where  $\beta_3$  is the coefficient of interest as it captures the interaction between time and treatment, that is, how the treatment affects post-reform labor supply as compared to the control group. The dependent variable *Labor Supply<sub>it</sub>* denotes the relative amount of labor, i.e., the amount of tasks solved over the amount of tasks to solve, subject  $i$  chooses to provide in every period.  $Post_t$  is a dummy variable that takes the value of 0 in pre-reform periods and 1 in post-reform periods; and  $Treat_i$  is a dummy variable indicating whether a subject is in the respective treatment *Prog,Flat25*, *Prog,Flat50*, *Flat25,Prog*, or *Flat50,Prog*. Additionally, we control for self-reported sociodemographic characteristics, such as age, being female, experience on MTurk, being full-time employed, having a household income below the US median and having obtained at least a bachelor's degree. Last, we use robust standard errors in all analyses and cluster them on the subject level.

First, we examine differences in pre- versus post-reform labor supply between each treatment group (i.e., tax reform treatments) and the control group (the baseline group). Estimation results are presented in Models (1) to (4) in Table 6. Only Model (1) shows a significant coefficient of interest at the  $p < 0.001$  level, implying that the introduction of a flat tax system with a relatively low tax rate (*Prog,Flat25*) significantly increases labor supply compared to no reform (*Prog,Prog*). Labor supply is eight percentage points higher than before the reform, compared to the control group. The relevant coefficients do not reach significance at conventional levels in the other models, suggesting that the other reforms did not affect labor supply relative to the baseline setting.

Therefore, our difference-in-differences estimation partly confirms that changes in the income tax rate affect labor provision (*H1*). We predicted that reform responses could either be driven by an income effect (*H2a*) or a substitution effect (*H2b*). We do not find evidence for the income effect. The descriptive results, presented above, suggest a substitution effect implying that an increase (decrease) in taxes reduces (increases) labor supply for the reforms *Prog,Flat25* and *Flat25,Prog*. However, the formal analysis only confirms the effect of a tax-decreasing reform *Prog,Flat25*.

*Result 1: Labor provision responds to a decrease, but not to an increase in the income*



tax rate.

*Result 2: A decrease in the income tax rate increases labor supply, suggesting a substitution effect. However, we find no evidence of a substitution effect in response to other tax reforms.*

	Diff-in-Diff – Labor Supply					
	(1)	(2)	(3)	(4)	(5)	(6)
Post Reform	-0.007 (0.006)	-0.007 (0.006)	-0.007 (0.006)	-0.007 (0.006)	-0.029* (0.012)	0.021 (0.015)
Prog,Flat25	-0.052* (0.022)				-0.034 (0.028)	
Prog,Flat50		-0.058** (0.021)				
Flat25,Prog			-0.026 (0.024)			
Flat50,Prog				-0.048 (0.025)		0.016 (0.029)
Post $\times$ Treat	0.083*** (0.009)	0.027 (0.016)	-0.022 (0.013)	0.022 (0.017)	0.106*** (0.014)	-0.005 (0.023)
Constant	0.947*** (0.055)	0.920*** (0.082)	0.950*** (0.067)	0.895*** (0.060)	1.003*** (0.056)	0.901*** (0.074)
Observations	3344	3216	3296	3392	3376	3344
R <sup>2</sup>	0.035	0.025	0.044	0.021	0.076	0.009
Controls	✓	✓	✓	✓	✓	✓

Table 6: Regression-DD estimates for reform effects

Notes: OLS estimates. The dependent variable is labor supply (in %). The reference category for Models (1)-(4) is the control group (*Prog,Prog*). The reference category for Model (5) is *Flat25,Prog*, and the reference category for Model (6) is *Prog,Flat50*. Socioeconomic controls are age, female, having obtained at least a bachelor's degree, full-time employment, below median household income, and MTurk experience. Table 10 in Appendix A displays controls and includes further individual characteristics such as risk aversion, loss aversion or understanding.

Robust standard errors clustered on the subject level are in parentheses.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05.

Models (5) and (6) include further robustness checks, holding the monetary incentives between treatments constant and checking whether the order of tax regimes matters. Model (5) estimates reform effects of *Prog,Flat25* using *Flat25,Prog* as the reference group. It confirms the estimation results of Model (1), underlining that the tax decrease in *Prog,Flat25* significantly increases labor supply by 10.6 percentage points at the  $p < 0.001$  level. This confirms a substitution effect in this treatment. Model (6) again shows that the tax-decreasing reform in *Flat50,Prog* does not render a coefficient that reaches conventional significance levels when using the tax-increasing reform *Prog,Flat50* as the monetarily equivalent reference group.

*Result 3: The order of tax regimes matters for reform responses. Only experiencing a tax decrease in a reform from a progressive to a flat tax regime increases labor supply.*

We can conclude from this section that labor supply responses to changes in the tax rate are asymmetrical. A tax decrease leads to more labor supply in *Prog,Flat25*. In this case, subjects can maximize their income by supplying more labor. However, a tax increase does not affect labor supply in *Prog,Flat50*. In this case, subjects can keep but not increase their income by providing more labor after the reform. A tax increase does also not affect labor supply even when subjects could optimize their post-reform income by providing less labor (*Flat25,Prog*). Finally, we find that a tax decrease does not affect labor supply, even when providing less labor post-reform would result in equivalent levels of income while at the same time increasing leisure (*Flat50,Prog*).

The results are striking as we would expect that the reform effects of *Prog,Flat25* and *Flat25,Prog* are identical. However, we do not find that subjects reduce labor supply in *Flat25,Prog* in order to optimize the after-tax income with minimal effort. This suggests that the possibility of increasing after-tax income provides a stronger incentive to increase labor supply than the option of decreasing labor supply and maintaining the maximum possible after-tax income. In other words, obtaining another 200 cents after-tax income is valued more than the time of working on 10 labor units.

The next question to address then is whether reform affectedness plays a role in these asymmetrical responses to tax reform. To investigate this issue, we split the sample in two groups according to pre-reform labor supply below or equal to 20 labor units and above 20 units. As discussed in Section 3 above, we expect that

labor supply responses to the reforms depend on the nature of the reform, individual affectedness, and individual preferences for leisure versus income (*H2c*). Regarding tax increases, we expect that affected individuals, i.e., individuals who provide more than 20 labor units in the pre-reform periods reduce their post-reform labor supply in *Flat25,Prog* (specific substitution effect). In *Prog,Flat50*, where the reform affects individuals who provide 20 labor units or less in the first half of the study, we expect that individuals increase their labor supply post-reform (specific income effect). Regarding tax decreases, we assume that behavioral responses depend on preferences for income and labor. Affected individuals in *Prog,Flat25* provide more than 20 labor units pre-reform, while in *Flat50,Prog* affected individuals provide 20 labor units or less before the reform. In particular, we anticipate that individuals who prefer income over leisure, *I-Types*, increase their labor supply when taxes decrease (specific substitution effect), while individuals who prefer leisure over income, *L-Types*, reduce their labor provision in this case (specific income effect).

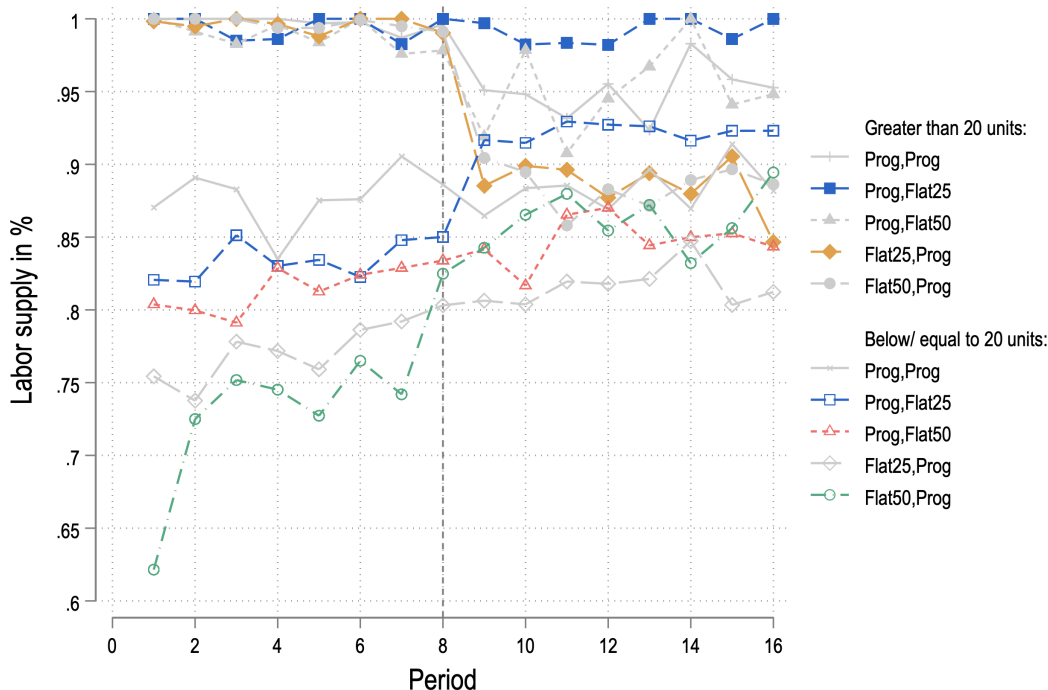


Figure 5: Evolution of labor supply by treatment and reform affectedness

Notes: This figure plots mean labor supply (in %) by treatment over 16 decision periods. The tax reform occurs after period eight (dashed line). In grey are treatment/ labor supply in *T0* combinations for which we predict no change, see Section 3.

Figure 5 plots how labor supply develops over time by treatment and reform affectedness, i.e., pre-reform labor supply ( $\leq$  or  $>$  20 units). We focus on groups for which we expect a change in labor supply. For individuals who provide low levels ( $\leq$  20) of labor in pre-reform periods, we find increases in post-reform labor supply in *Prog,Flat25* and *Flat50,Prog*. This suggests that in both treatments, the *I-Types*, who aim to increase their  $T1$  income when taxes decrease, dominate (specific substitution effect). In *Prog,Flat50*, there is a slight increase in labor supply, which is in line with our expectation that both *I-Types* and *L-Types* increase labor after the reform (specific income effect).

Examining the effect of tax reform on individuals who provide high levels ( $>$  20) of labor in the pre-reform periods, we find that there is no noticeable change in labor supply in *Prog,Flat25*. This is plausible since this group already chooses labor supply at nearly full capacity pre-reform and thus cannot increase it further. For *Flat25,Prog*, we observe a strong decrease in post-reform labor supply as individuals experience a tax increase. This is in line with our expectation that pre-reform income levels are unobtainable after the reform and therefore both *I-Types* and *L-Types* decrease labor supply (specific substitution effect).

We next investigate whether these observations hold when we add socio-demographic variables and account for trends in the control condition. Table 7 displays the results of difference-in-differences estimations splitting the sample by reform affectedness according to pre-reform labor supply decisions. The findings confirm the descriptive intuition. Model (1) shows a positive and significant coefficient of interest in *Prog,Flat25* for individual decisions providing 20 labor units or less in the pre-reform periods. The reform leads to a labor supply increase by 8.3 percentage points. Model (2) shows that for pre-reform decisions of  $>$  20 labor units in *Prog,Flat25*, there is also an increase in labor although at only half of the effect size with 4.6 percentage points. Both results are in line with our predictions for *I-Types* who increase labor to maximize income after a reform, indicating a specific substitution effect.

Models (3) and (4) show that neither for decisions less nor more than 20 labor units, there is a change in labor supply post-reform in *Prog,Flat50* compared to the control condition. This is partly in line with our predictions. Specifically, we predicted that both types of individuals (*I-Types* and *L-Types*) who provided less than 20 labor units pre-reform would increase their labor supply in response to the tax increase in order to maximize, or maintain, their post-reform income. In contrast, for individual decisions of  $>$  20 labor units pre-reform, we predicted no change in

labor supply, as the tax rate does not change.

In contrast to our predictions, Model (5) reveals a moderate and marginally significant increase in post-reform labor supply in *Flat25,Prog* among decisions of 20 or less labor units in the pre-reform periods. This finding is surprising, because the reform does not introduce a change in the tax rate for this group. For decisions of more than 20 labor units in the pre-reform periods of this treatment, we find a significant decrease of 6.2 percentage points post-reform in Model (6). This confirms our expectation that the tax increase for such decisions should reduce labor provision, irrespective of whether individuals value income or labor more (specific substitution effect).

	Diff-in-Diff – Labor Supply							
	$\leq 20$	$> 20$	$\leq 20$	$> 20$	$\leq 20$	$> 20$	$\leq 20$	$> 20$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post Reform	0.005 (0.005)	-0.049** (0.015)	0.005 (0.005)	-0.049** (0.015)	0.005 (0.005)	-0.049** (0.015)	0.005 (0.005)	-0.049** (0.015)
Prog,Flat25	-0.046 (0.025)	-0.003 (0.005)						
Prog,Flat50			-0.059* (0.023)	-0.012* (0.006)				
Flat25,Prog					-0.098** (0.032)	0.001 (0.004)		
Flat50,Prog							-0.139*** (0.035)	-0.003 (0.003)
Post $\times$ Treat	0.083*** (0.009)	0.046** (0.015)	0.028 (0.018)	0.011 (0.023)	0.039* (0.017)	-0.062*** (0.018)	0.119*** (0.025)	-0.062*** (0.017)
Constant	0.939*** (0.065)	1.002*** (0.014)	0.878*** (0.096)	1.011*** (0.021)	0.959*** (0.090)	0.955*** (0.021)	0.885*** (0.079)	1.000*** (0.020)
Observations	2786	558	2592	624	2156	1140	2224	1168
R <sup>2</sup>	0.041	0.119	0.029	0.089	0.075	0.214	0.068	0.245
Controls	✓	✓	✓	✓	✓	✓	✓	✓

Table 7: Regression-DD estimates by reform affectedness

Notes: OLS estimates. The dependent variable is labor supply (in %). The reference category for all Models is the control group (*Prog,Prog*). For each treatment, we estimate two models in which we restrict the sample to labor supply decisions of either  $\leq 20$  or  $> 20$  labor units in *T0*. Control variables include age, female, bachelor's degree, full-time employment, below median household income and MTurk experience.

Table 11 in Appendix A displays all controls, adding loss aversion and other variables.

Robust standard errors clustered on the subject level are in parentheses.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

In Model (7), the relevant coefficient is positive and significant, indicating a

11.9 percentage point increase in labor supply for pre-reform decisions  $\leq 20$  units in *Flat50,Prog*. This finding again indicates dominance of *I – Type* individuals in our sample, who prefer income over leisure. The response to the tax decrease thus suggests a specific substitution effect. Regarding  $> 20$  units pre-reform decisions, we find a significant 6.2 percentage points decrease in labor supply in this treatment. This results is again unexpected, because the tax reform does not introduce a change of the tax rate for these individuals. Taken together, the heterogeneity analysis confirms *H2c*: behavioral responses are dependent on the type of reform, reform affectedness and preferences for leisure versus income.

*Result 4: Labor supply responses to tax policy reform are mostly driven by substitution effects in decisions which are affected by the reform.*

#### 4.4 Further margins of reform responses and bunching

In addition to defining labor supply as the share of labor units that individuals actually provide relative to the amount of labor that they may potentially provide, we examine the extensive and intensive margins of adjustment to the reforms. The two margins allow us to disentangle whether a higher fraction of responsive subjects or a higher/ lower amount of supplied labor characterizes different reform responses between treatments. We split the extensive margin into three dummy variables, indicating whether a taxpayer i) does not respond to a reform, ii) reduces labor due to the reform, or iii) increases labor due to the reform. For the intensive margin, we focus on how much labor is provided conditional on the taxpayer not changing, reducing or increasing labor supply after the reform.

Figure 6 plots average reform responses for the extensive and intensive margins by treatment. In Panel A in the control condition without a reform (*Prog,Prog*), we find that 87% of labor supply decisions do not change in periods 9-16 compared to periods 1-8. In the reform treatments, we find strong and significant differences regarding decisions that increase or reduce labor supply after the reform. In *Prog,Flat25*, 49% of decisions increase labor supply in post-reform periods. This fraction amounts to 37% in *Prog,Flat50*, and only to 12% and 17% in *Flat25,Prog* and *Flat50,Prog*, respectively. Instead, the fraction of decisions to reduce labor supply is more pronounced in these treatments. In *Flat25,Prog*, it is at 34% and in *Flat50,Prog* at 32%. For all the reform treatments, we can observe that the fraction of decisions that do not

respond to the reform ranges between 47% and 54%, which is significantly smaller than in the control condition.

Regression results in Table 12 in Appendix A confirm significant differences in the decisions to change labor supply between the control condition and the reform treatments on the extensive margin. All reforms lead to a significantly higher fraction of decisions to increase labor supply as compared to the control condition, with this effect being strongest in *Prog,Flat25*, as expected. Regarding decisions to reduce labor supply after the reform, the estimates confirm that *Flat25,Prog* and *Flat50,Prog* lead to significantly more reduction behavior, also when including control variables.

This analysis of different labor supply responses on the extensive margin explains why we find significant differences between *Prog,Prog* and *Prog,Flat25*, but not for any of the other treatments in the earlier analysis. The fraction of labor supply increases due to the reform is the largest of any of the treatments. However, labor supply-reducing responses are the least frequent in *Prog,Flat25*. In all the other reforms, we also find significantly different reform responses. However, when subsuming responses into the relative provision of labor supply, increasing and reducing responses may cancel each other out, depending on how strongly subjects respond on the intensive margin.

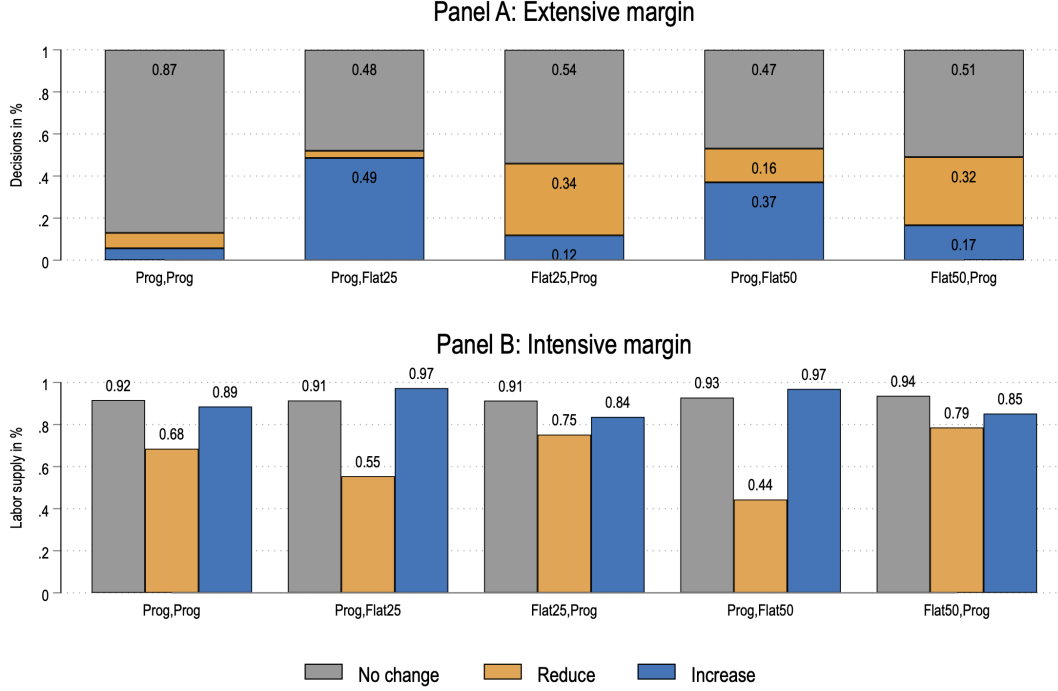


Figure 6: Average reform responses on the extensive and intensive margin by treatment

Notes: Panel A shows stacked bars of average behavioral responses to reforms by treatment. No change means that taxpayers do not change their labor supply between pre-and post-reform periods. Reduce implies that taxpayers choose less labor after the reform, whereas increase means that they choose more labor after the reform. Panel B shows the labor supply (in %), conditional on having reduced or increased labor after the reform or not having changed labor.

In Panel B in Figure 6, we display labor supply conditional on an increase, decrease or no change response by treatment. It is evident that decisions to increase labor supply after a reform to a flat tax system (*Prog,Flat25* and *Prog,Flat50*) imply choosing nearly the maximum possible number of tasks in the post-reform period, with an average labor supply of 97%. In the other reform treatments, we observe that the fraction of decisions to increase labor supply is low with 12% and 17% on the extensive margin. However, when subjects decide to increase their labor supply after the reform, they provide 85% of the maximum labor supply possible. The intensive margin for decisions to increase labor supply does not significantly differ between *Flat25,Prog* and *Flat50,Prog*. Interestingly, however, the intensive margin of decisions to reduce labor supply in treatments *Flat25,Prog* and *Flat50,Prog* is very similar to that of decisions to increase labor supply. In *Prog,Flat25* and *Prog,Flat50*,



decisions to provide less labor imply significantly lower labor supply with 55% and 44%, respectively.

Additional analyses including control variables, presented in Table 13 in Appendix A corroborate that decisions to increase labor supply in *Prog,Flat25* and *Prog,Flat50* result in significantly higher labor supply than in the control group. Similarly, labor supply of reduce decisions is significantly lower in *Prog,Flat50*, which also confirms the descriptive results.

Overall, these findings indicate a meaningful labor supply increase on both the extensive and intensive margin in *Prog,Flat25* after the reform, compared to the other reforms and the control condition. In the other treatments, reform responses cancel each other out on the aggregate, as they are mixed on the extensive margin and similarly pronounced in both directions on the intensive margin.

To elaborate whether subjects rationally adapt to progressive tax systems, we further decompose the margins of reform responses and analyze bunching behavior. There is a vast literature that estimates bunching behavior at discontinuities in different contexts using administrative data (see Kleven, 2016, for an overview). In our setting, we expect bunching to occur at the cutoff of 20 labor units, or a gross income of 400 cents.<sup>7</sup>

The share of bunching decisions is by far the highest in *Prog,Prog*, with around 60%. This is in line with our assumptions, as subjects face the progressive tax regime in all 16 periods. In the progressive tax regime, the incentive to bunch at 20 labor units is particularly pronounced, since the average tax rate jumps by 25 percentage points when subjects choose to provide more than 20 labor units.

Figure 7 plots pre- and post-reform bunching behavior by treatment. We see significantly higher but also lower fractions of bunching in the reform treatments, with 68% of decisions pre-reform in *Prog,Flat25*, 48% of decisions in *Flat25,Prog* post-reform, 52% in *Prog,Flat50* pre-reform, and 51% in *Flat50,Prog* after the reform. The highest fraction of bunching decisions is in *Prog,Flat25*, suggesting that taxpayers optimized their labor supply more than in any of the other treatment.

The results in Table 14 in Appendix A confirm that all reforms lead to significantly different bunching responses compared to the control condition. Confirming the descriptive results, *Prog,Flat25* displays the strongest decline in bunching, estimated at 45.2 percentage points. It is evident that when flat tax regimes change into

---

<sup>7</sup>To identify bunching, we exclude decisions where the maximum possible labor supply is below or equal to the threshold of 20 labor units, or a gross income of 400 cents, because with these incomes, bunching at the threshold is not possible. As a consequence, we limit the sample to ten observations per subject in this analysis.

progressive tax regimes, the difference in bunching is much lower, around half the size, than for the reverse type of reform. In other words: Subjects can optimize labor supply better when they experience a reform from a progressive tax regime to a flat tax regime.

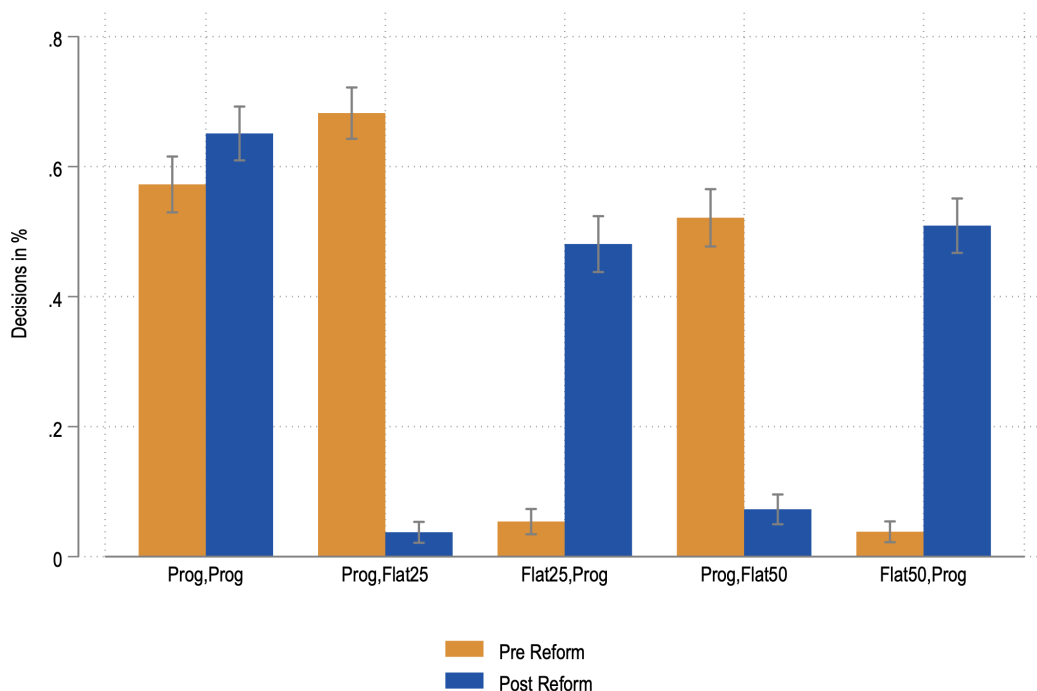


Figure 7: Bunching behavior by treatment

Notes: This figure shows average proportions of bunching decisions with 95% confidence intervals by treatment pre- and post-reform. We display a dummy variable denoting whether a taxpayer chooses 20 tasks in the periods where the maximum number of tasks possible is above 20. Pre-reform periods are periods 1 to 8, and post-reform are periods 9 to 16.

## 4.5 Mechanisms to explain reform responses

Our analysis of the labor provision responses to tax reform has yielded several important insights. Tax reforms affect labor supply, but only when they entail a tax decrease from a progressive to a flat tax regime. Positive labor supply responses in this reform are thus dominated by a substitution effect. Furthermore, our sample is characterized by *I – Type* individuals who aim to increase their income post-reform when affected by the reform. However, these results also raise important questions: What mechanisms could explain the strong response to the reform that reduces taxes,

changing from a progressive to a flat tax regime? And what drives the lack of optimal adaptation to progressive tax regimes? One answer is that subjects reduce their labor provision below the threshold of 20. This ‘overbunching’ behavior occurs in all progressive tax regimes, with a fraction of around 10% of subjects reducing labor more than optimal. Only in *Prog,Prog* and *Prog,Flat25*, it is not so pronounced, with 3% to 8% of subjects adapting more than theoretically expected. This could be due to misperceptions, which we examine in more detail next. We explore two avenues for behavioral explanations, first, whether fairness perceptions affect reform responses. Second, to what extent misperceptions of the tax regime explain reform responses.

Tax reforms often entail fairness discussions. Progressive tax systems are typically deemed fairer, however, public opinion is sensitive as to how the progressive tax system is presented (e.g., Pántya et al., 2016; Reimers, 2009; Roberts et al., 1994). At the same time, complex tax regimes are perceived as more unfair (Benzarti & Wallossek, 2024). Moreover, several studies show that taxpayers feel they were unfairly treated when facing high tax rates and hence reduce labor supply to punish the tax authorities (Keser et al., 2020; Lévy-Garboua et al., 2009; Hundsdoerfer & Matthaei, 2022).

We use a dummy variable, indicating whether subjects consider the tax burden as very fair or somewhat fair.<sup>8</sup> In general, 58.4% of subjects perceive the tax burden as fair. Half of the subjects in *Prog,Prog* indicate that the tax payment was fair. In the reform treatments, fairness perceptions are higher. Comparing fairness perceptions by treatment, we find that 71% of taxpayers perceive the tax payment in *Prog,Flat25* as fair, whereas the other reforms are perceived as significantly less fair. This explains the asymmetry and why we do not find evidence for a substitution effect in *Flat25,Prog* too. It suggests that subjects are more willing to increase labor, if they feel that a tax reform is fair.

To assess whether misperceptions of the tax regime affect reform responses (or the lack thereof), we examine whether the calculation of the tax payment in the progressive regime varies with the type of reform.

In the post-experimental questionnaire, we ask the subjects to indicate how much

---

<sup>8</sup>We ask subjects about their fairness perceptions on a 5-point Likert scale. The question is: “If you think about choosing an income in the experiment, would you say that the tax payment was fair or unfair?”.

tax they need to pay in the progressive tax regime.<sup>9</sup> According to this simple question, on average, 81.2% of the subjects correctly understand the effect of progressive tax rates on after-tax income. The fraction of subjects that chooses the correct tax payment is higher in *Prog,Prog* with 86.3% than in the other treatments. This is not surprising since subjects are exposed to the progressive tax regime for 16 instead of just eight periods. Compared to the other reform treatments, understanding is slightly higher in the reform *Prog,Flat25*.

It is interesting to see that 13.0% subjects in our sample miscalculate the tax payment post-experimentally. They misperceive the threshold at 20 labor units, or 400 cents, as a tax kink instead of a notch, and hence believe that only the income above the threshold is taxed at 50%. This is surprising since the instructions on the threshold are very salient and include an example. This might explain why reform responses are less pronounced in *Prog,Flat50* and *Flat50,Prog*, as 14.1% and 16.4% of subjects mistake the notch for a kink, respectively.

## 5 Conclusion

Discontinuous changes in the tax burden such as tax notches are a common instrument to induce progressivity in tax systems. To study labor responses to tax reforms that discontinuously affect the tax burden, we conduct an online labor-market experiment with 522 subjects.

Our findings indicate asymmetric responses to tax reform as well substantial heterogeneity at the individual level. In particular, we find an increase in labor supply in response to a tax reform only when the reform reduces the tax burden at the upper end of the income distribution by transforming a progressive into a flat tax regime. In contrast, we do not find evidence of labor supply effects of tax reforms increasing the tax burden nor do we find effects of reforms resulting in a decrease of the tax burden due to the transformation of a flat tax into a progressive tax regime. Moreover, we find evidence of path dependency in behavioral responses to tax reform. In particular, we find that post-reform labor responses hinge on the design of the pre-reform tax system. We also find substantial heterogeneity in labor supply effects at the individual level, with aggregate effects driven mostly by individuals who are

---

<sup>9</sup>The question is phrased like this: Suppose you chose to earn an income of 420 cents. The tax rate is 25% for incomes equal to or below 400 cents. The tax rate is 50% on the entire income if the income exceeds 400 cents. What would be the tax payment? [105 cents, 110 cents, 210 cents].

affected by the reform. Finally, a post-experimental survey reveals that subjects are more likely to optimize labor supply, consider the reform fairer, and compute the tax burden more accurately when the tax reform decreases the tax burden by transforming a progressive into a flat tax regime.

Understanding the mechanisms underlying labor supply responses to tax reforms is highly policy relevant. We find that the type of reform has asymmetric effects on labor supply and responses to tax policy reform are not necessarily in line with theoretical predictions, even in a world without frictions. In order to achieve the intended effects of a tax reform, policymakers should thus not only focus on the design of the reform, but also on behavioral aspects, such as path dependency and fairness concerns.

## References

- Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2), 470–492.
- Abeler, J., & Jäger, S. (2015). Complex Tax Incentives. *American Economic Journal: Economic Policy*, 7(3), 1–28. doi: 10.1257/pol.20130137.
- Agarwal, S., Liu, C., & Souleles, N. S. (2007). The Reaction of Consumer Spending and Debt to Tax Rebates—Evidence from Consumer Credit Data. *Journal of Political Economy*, 115(6), 986–1019.
- Amberger, H. J., Eberhartinger, E., & Kasper, M. (2023). Tax-Rate Biases in Tax Decisions: Experimental Evidence. *Journal of the American Taxation Association*, 45(1), 7–34.
- Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting Interactive Experiments Online. *Experimental Economics*, 21(1), 99–131. doi: <https://doi.org/10.1007/s10683-017-9527-2>.
- Augenblick, N., Niederle, M., & Sprenger, C. (2015). Working Over Time: Dynamic Inconsistency in Real Effort Tasks. *The Quarterly Journal of Economics*, 130(3), 1067–1115. doi: <https://doi.org/10.1093/qje/qjv020>.
- Avram, S. (2015). *Benefit Losses Loom Larger than Taxes: The Effects of Framing and Loss Aversion on Behavioural Responses to Taxes and Benefits*. ISER Working Paper Series. Retrieved from <https://www.iser.essex.ac.uk/wp-content/uploads/files/working-papers/iser/2015-17.pdf>
- Azmat, G. (2014). Evaluating the Effectiveness of In-Work Tax Credits. *Empirical Economics*, 46, 397–425. doi: <https://doi.org/10.1007/s00181-012-0678-x>.
- Bachas, P., & Soto, M. (2018). *Not(ch) Your Average Tax System: Corporate Taxation under Weak Enforcement*. World Bank Policy Research Working Paper No. 8524. Retrieved from <https://ssrn.com/abstract=3238367>
- Bastian, J., & Lochner, L. (2022). The Earned Income Tax Credit and Maternal Time Use: More Time Working and Less Time with Kids? *Journal of Labor Economics*, 40(3), 573–611. doi: <https://doi.org/10.1086/717729>.

- Benzarti, Y., & Wallossek, L. (2024). Rising Income Tax Complexity. *National Tax Journal*, 77(1), 143–173.
- Blaufus, K., Chirvi, M., Huber, H.-P., Maiterth, R., & Sureth-Sloane, C. (2022). Tax Misperception and its Effects on Decision Making—Literature Review and Behavioral Taxpayer Response Model. *European Accounting Review*, 31(1), 111–144. doi: <https://doi.org/10.1080/09638180.2020.1852095>.
- Blinder, A. S., & Rosen, H. S. (1985). Notches. *American Economic Review*, 75(4), 736–747.
- Blumkin, T., Ruffle, B. J., & Ganun, Y. (2012). Are Income and Consumption Taxes Ever Really Equivalent? Evidence from a Real-Effort Experiment with Real Goods. *European Economic Review*, 1200–1219. doi: <https://doi.org/10.1016/j.euroecorev.2012.06.001>.
- Chetty, R. (2012). Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply. *Econometrica*, 80(3), 969–1018. doi: <https://doi.org/10.3982/ECTA9043>.
- Chetty, R., Friedman, J. N., & Saez, E. (2013). Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings. *American Economic Review*, 103(7), 2683–2721. doi: 10.1257/aer.103.7.2683.
- Chetty, R., & Saez, E. (2013). Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients. *American Economic Journal: Applied Economics*, 5(1), 1–31. doi: 10.1257/app.5.1.1.
- Corgnet, B., Hernán-González, R., & Schniter, E. (2015). Why Real Leisure Really Matters: Incentive Effects on Real Effort in the Laboratory. *Experimental Economics*, 18(2), 284–301. doi: <https://doi.org/10.1007/s10683-014-9401-4>.
- Cuskley, C., & Sulik, J. (2022). *The Burden for High-Quality Online Data Collection Lies With Researchers, Not Recruitment Platforms*. PsyArXiv. doi: 10.31234/osf.io/w7qy9.
- Dickinson, D. L. (1999). An Experimental Examination of Labor Supply and Work Intensities. *Journal of Labor Economics*, 17(4), 638–670. doi: <https://doi.org/10.1086/209934>.

- Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and Dynamics of Mechanical Turk Workers. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (pp. 135–143). doi: <https://doi.org/10.1145/3159652.3159661>.
- Di Porto, E., Garibaldi, P., Mastrobuoni, G., & Naticchioni, P. (2022). *The Perverse Effect of Flexible Work Arrangements on Informality*. Available at SSRN 4295968. doi: <http://dx.doi.org/10.2139/ssrn.4295968>.
- Doerrenberg, P., & Duncan, D. (2014). Experimental Evidence on the Relationship between Tax Evasion Opportunities and Labor Supply. *European Economic Review*, 68, 48–70. doi: <https://doi.org/10.1016/j.euroecorev.2014.02.005>.
- Doerrenberg, P., Duncan, D., & Löffler, M. (2023). Asymmetric labor-supply responses to wage changes: Experimental evidence from an online labor market. *Labour Economics*, 81, 102305.
- Dolado, J. J., Lalé, E., & Turon, H. (2022). *Zero-hours Contracts in a Frictional Labor Market*. Available at SSRN 3988202. doi: <http://dx.doi.org/10.2139/ssrn.3988202>.
- Eriksson, T., Poulsen, A., & Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6), 679–688.
- Fochmann, M., & Weimann, J. (2013). The Effects of Tax Salience and Tax Experience on Individual Work Efforts in a Framed Field Experiment. *FinanzArchiv/Public Finance Analysis*, 511–542. doi: 10.1628/001522113X675692.
- Fochmann, M., Weimann, J., Blaufus, K., Hundsdoerfer, J., & Kiesewetter, D. (2013). Net Wage Illusion in a Real-Effort Experiment. *The Scandinavian Journal of Economics*, 115(2), 476–484. doi: <https://doi.org/10.1111/sjoe.12007>.
- Giamattei, M., Yahosseini, K. S., Gächter, S., & Molleman, L. (2020). LI-ONESS Lab: A Free Web-Based Platform for Conducting Interactive Experiments Online. *Journal of the Economic Science Association*, 6(1), 95–111. doi: <https://doi.org/10.1007/s40881-020-00087-0>.
- Gibson, J., Norton, D. A., & White, R. A. (2019). The Backward Hustle: An Experimental Investigation of Tax Code Notches and Labor Sup-



- ply. *Journal of Economic Behavior and Organization*, 166, 432–445. doi: <https://doi.org/10.1016/j.jebo.2019.07.011>.
- Hauser, D. J., Moss, A. J., Rosenzweig, C., Jaffe, S. N., Robinson, J., & Litman, L. (2022). Evaluating CloudResearch’s Approved Group as a Solution for Problematic Data Quality on MTurk. *Behavior Research Methods*, 1–12. doi: <https://doi.org/10.3758/s13428-022-01999-x>.
- Hayashi, A. T., Nakamura, B. K., & Gamage, D. (2013). Experimental Evidence of Tax Salience and the Labor–Leisure Decision: Anchoring, Tax Aversion, or Complexity? *Public Finance Review*, 41(2), 203–226. doi: 10.1177/1091142112460726.
- He, T.-S. (2020). The framing effect of tax–transfer systems. *Journal of the Economic Science Association*, 6(2), 213–225. doi: <https://doi.org/10.1007/s40881-020-00095-0>.
- Hundsdoerfer, J., & Matthaei, E. (2022). Gender Discriminatory Taxes, Fairness Perception, and Labor Supply. *FinanzArchiv*, 78(1-2), 156–207. doi: 10.1628/fa-2021-0020.
- Johnson, D. S., Parker, J. A., & Souleles, N. S. (2006). Household Expenditure and the Income Tax Rebates of 2001. *American Economic Review*, 96(5), 1589–1610.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 363–391.
- Keser, C., Masclet, D., & Montmarquette, C. (2020). Labor Supply, Taxation, and the Use of Tax Revenues: A Real-Effort Experiment in Canada, France, and Germany. *Public Finance Review*, 48(6), 714–750. doi: <https://doi.org/10.1177/1091142120960491>.
- Kessler, J. B., & Norton, M. I. (2016). Tax Aversion in Labor Supply. *Journal of Economic Behavior and Organization*, 124, 15–28. doi: <https://doi.org/10.1016/j.jebo.2015.09.022>.
- Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, 8, 435–464. doi: <https://doi.org/10.1146/annurev-economics-080315-015234>.

- Kleven, H. J., & Waseem, M. (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *The Quarterly Journal of Economics*, 128(2), 669–723. doi: <https://doi.org/10.1093/qje/qjt004>.
- Kostøl, A. R., & Myhre, A. S. (2021). Labor Supply Responses to Learning the Tax and Benefit Schedule. *American Economic Review*, 111(11), 3733–66. doi: 10.1257/aer.20201877.
- Krekó, J., Prinz, D., & Weber, A. (2022). *Take-up and Labor Supply Responses to Disability Insurance Earnings Limits*. IZA Discussion Paper. Retrieved from <https://docs.iza.org/dp15377.pdf>
- Lévy-Garboua, L., Masclet, D., & Montmarquette, C. (2009). A Behavioral Laffer Curve: Emergence of a Social Norm of Fairness in a Real Effort Experiment. *Journal of Economic Psychology*, 30(2), 147–161. doi: <https://doi.org/10.1016/j.joep.2008.09.002>.
- Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A Versatile Crowdsourcing Data Acquisition Platform for the Behavioral Sciences. *Behavior Research Methods*, 49(2), 433–442. doi: 10.3758/s13428-016-0727-z.
- Liu, L., & Lockwood, B. (2015). VAT Notches. In *Proceedings. annual conference on taxation and minutes of the annual meeting of the national tax association* (Vol. 108, pp. 1–51).
- Liu, L., Lockwood, B., Almunia, M., & Tam, E. H. (2021). VAT Notches, Voluntary Registration, and Bunching: Theory and UK Evidence. *Review of Economics and Statistics*, 103(1), 151–164.
- Masclet, D., & Montmarquette, C. (2008). Experimental Approach to the Effect of Taxation on the Labor Supply: A Comparative Study of Tax Systems. *Economie & Prévision*, 182(1), 47–59. doi: <https://doi.org/10.3917/ecop.182.0047>.
- Mavrokonstantis, P., & Seibold, A. (2022). Bunching and Adjustment Costs: Evidence from Cypriot Tax Reforms. *Journal of Public Economics*, 214, 104727.
- Mian, A., & Sufi, A. (2012). The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program. *The Quarterly Journal of Economics*, 127(3), 1107–1142.

- Mohnen, A., Pokorny, K., & Sliwka, D. (2008). Transparency, inequity aversion, and the dynamics of peer pressure in teams: Theory and evidence. *Journal of Labor Economics*, 26(4), 693–720.
- Pántya, J., Kovács, J., Kogler, C., & Kirchler, E. (2016). Work Performance and Tax Compliance in Flat and Progressive Tax Systems. *Journal of Economic Psychology*, 56, 262–273. doi: <https://doi.org/10.1016/j.joep.2016.08.002>.
- Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a Sufficient Condition for Data Quality on Amazon Mechanical Turk. *Behavior Research Methods*, 46(4), 1023–1031. doi: <https://doi.org/10.3758/s13428-013-0434-y>.
- Prissé, B., & Jorrat, D. (2022). Lab vs Online Experiments: No Differences. *Journal of Behavioral and Experimental Economics*, 100, 101910. doi: <https://doi.org/10.1016/j.socec.2022.101910>.
- Reimers, S. (2009). A Paycheck Half-Empty or Half-Full? Framing, Fairness and Progressive Taxation. *Judgment and Decision Making*, 4(6), 461–466.
- Roberts, M. L., Hite, P. A., & Bradley, C. F. (1994). Understanding Attitudes Toward Progressive Taxation. *Public Opinion Quarterly*, 58(2), 165–190.
- Saving, J. L., & Viard, A. D. (2021). Notches in the Tax System: The Good, the Bad, and the Ugly. *Tax Notes Federal*, 171(8), 1253–1263. Retrieved from <https://www.aei.org/wp-content/uploads/2021/06/Saving-and-Viard-On-the-Margin-May-24-2021.pdf?x91208>
- Scarfe, R. (2019). *Flexibility or Certainty? The Aggregate Effects of Casual Jobs on Labour Markets* (Vol. 294). Retrieved from [http://www.econ.ed.ac.uk/papers/id294\\_esedps.pdf](http://www.econ.ed.ac.uk/papers/id294_esedps.pdf)
- Schneider, C., & Mill, W. (2022). *The Bright Side of Tax Evasion*. Available at SSRN. doi: <http://dx.doi.org/10.2139/ssrn.4024275>.
- Slemrod, J. (2013). Buenas Notches: Lines and Notches in Tax System Design. *eJournal of Tax Research*, 11, 259. Retrieved from <https://heinonline.org/HOL/P?h=hein.journals/ejotaxrs11&i=267>

- Soldatos, G. T. (2021). Intertemporal Substitution in Labour Supply and Tax Revenue Use. *Journal of Public Finance and Public Choice*, 36(2), 209–231. doi: <https://doi.org/10.1332/251569121X16144127668408>.
- Steiner, V., & Wakolbinger, F. (2013). Wage Subsidies, Work Incentives, and the Reform of the Austrian Welfare System. *Empirica*, 40, 259–285. doi: <https://doi.org/10.1007/s10663-012-9191-x>.
- 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. doi: <https://doi.org/10.1016/j.jpubeco.2020.104162>.
- Webb, M. A., & Tangney, J. P. (2022). Too Good to Be True: Bots and Bad Data From Mechanical Turk. *Perspectives on Psychological Science*, 17456916221120027. doi: <https://doi.org/10.1177/1745691622112002>.

## A Additional analyses

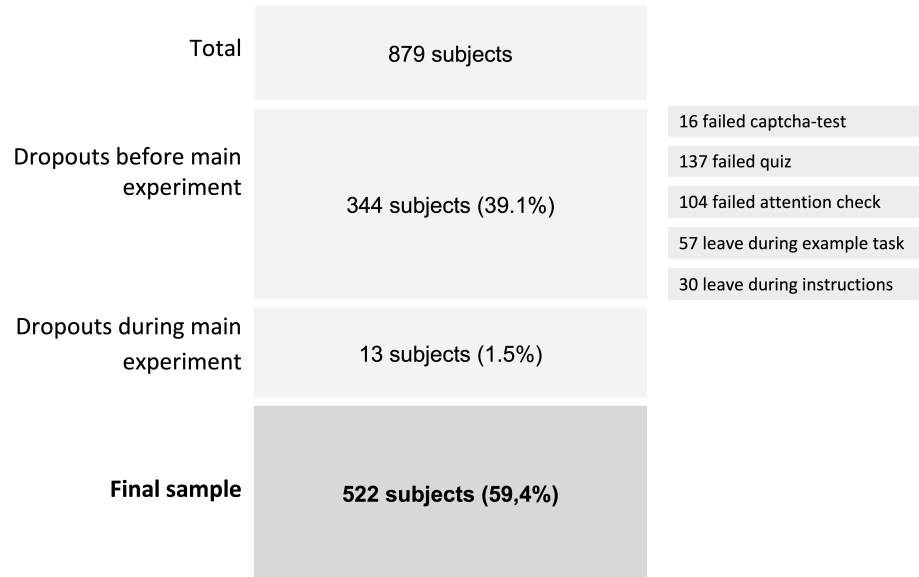


Figure 8: Dropouts at different stages of the experiment

	Probit (ME) – Dropouts
Age	-0.00225 (-1.80)
Female	0.00417 (0.12)
At Least Bachelor’s Degree	0.105** (2.98)
Full-time Employment	-0.0407 (-1.16)
Below Median Income	-0.0153 (-0.42)
MTurk Experience	0.000365 (0.25)
Observations	842

Table 8: Attrition analysis

Notes: The dependent variable is a dummy variable indicating whether an individual dropped out of the experiment. Independent variables are sociodemographic characteristics surveyed before the experiment. Note that this analysis does not include observations from individuals that dropped out before answering the sociodemographic questionnaire. Coefficients are marginal effects; standard errors are in parentheses.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , and \*  $p < 0.05$ .

	Prog,Prog	Prog,Flat25	Flat25,Prog	Prog,Flat50	Flat50,Prog	Total
Age	42.92 (10.41)	40.96 (13.06)	43.46 (13.73)	40.40 (12.24)	42.93 (13.75)	42.15 (12.74)
Female	0.63 (0.49)	0.54 (0.50)	0.62 (0.49)	0.58 (0.50)	0.53 (0.50)	0.58 (0.49)
At Least Bachelor's Degree	0.65 (0.48)	0.60 (0.49)	0.52 (0.50)	0.62 (0.49)	0.61 (0.49)	0.60 (0.49)
Full-time Employment	0.50 (0.50)	0.45 (0.50)	0.45 (0.50)	0.55 (0.50)	0.51 (0.50)	0.49 (0.50)
Below Median Income	0.56 (0.50)	0.52 (0.50)	0.68 (0.47)	0.60 (0.49)	0.60 (0.49)	0.59 (0.49)
MTurk Experience	14.29 (11.16)	13.99 (10.57)	14.26 (11.70)	13.28 (12.10)	12.62 (11.43)	13.68 (11.37)

Table 9: Sociodemographic characteristics by treatment

Notes: Standard deviations in parentheses. Age in years, MTurk experience in hours, the other variables as dummies.

	Diff-in-Diff – Labor Supply					
	(1)	(2)	(3)	(4)	(5)	(6)
Post Reform	-0.007 (0.006)	-0.007 (0.006)	-0.007 (0.006)	-0.007 (0.006)	-0.029* (0.012)	0.021 (0.016)
Prog,Flat25	-0.048* (0.019)				-0.032 (0.028)	
Prog,Flat50		-0.062** (0.021)				
Flat25,Prog			-0.027 (0.023)			
Flat50,Prog				-0.045 (0.025)		0.016 (0.029)
Post × Treat	0.083*** (0.009)	0.027 (0.016)	-0.022 (0.013)	0.022 (0.017)	0.106*** (0.014)	-0.005 (0.023)
Age	-0.001 (0.001)	-0.000 (0.001)	-0.002 (0.001)	-0.000 (0.001)	-0.004** (0.001)	-0.001 (0.001)
Female	-0.015 (0.021)	-0.021 (0.027)	-0.003 (0.022)	0.006 (0.021)	0.020 (0.028)	0.005 (0.031)
Bachelor's Degree	0.023 (0.023)	0.017 (0.028)	0.043 (0.026)	-0.014 (0.025)	0.056* (0.027)	0.002 (0.032)
Full-time Employed	0.015 (0.023)	-0.023 (0.027)	-0.015 (0.028)	0.040 (0.025)	-0.028 (0.028)	-0.020 (0.032)
Below Median Income	0.013 (0.023)	-0.021 (0.026)	0.061* (0.028)	0.026 (0.025)	0.038 (0.028)	-0.015 (0.028)
MTurk Experience	-0.000 (0.001)	0.001 (0.001)	-0.000 (0.001)	-0.001 (0.001)	0.000 (0.001)	0.000 (0.001)
Regime Fair	-0.009 (0.019)	0.011 (0.024)	0.022 (0.022)	0.010 (0.021)	-0.008 (0.022)	-0.010 (0.026)
Task Easy	-0.028 (0.022)	-0.016 (0.025)	0.001 (0.025)	-0.010 (0.025)	-0.020 (0.030)	-0.015 (0.034)
Tax Understood	0.074 (0.049)	0.003 (0.036)	0.030 (0.039)	0.046 (0.037)	0.077 (0.047)	0.048 (0.039)
Financially Literate	0.030 (0.024)	-0.013 (0.026)	0.027 (0.025)	-0.000 (0.023)	0.054* (0.024)	-0.022 (0.029)
Risk Averse	-0.037 (0.020)	0.028 (0.024)	-0.006 (0.025)	0.008 (0.022)	-0.050 (0.027)	0.026 (0.031)
Loss Averse	-0.002 (0.027)	0.021 (0.030)	-0.016 (0.024)	0.012 (0.024)	-0.011 (0.030)	0.039 (0.034)
Tax Honest	0.061 (0.045)	-0.033 (0.034)	-0.023 (0.029)	0.005 (0.027)	0.067 (0.051)	-0.020 (0.034)
Conservative	-0.005 (0.021)	-0.021 (0.025)	0.001 (0.023)	-0.006 (0.024)	-0.003 (0.025)	-0.010 (0.032)
Constant	0.853*** (0.094)	0.939*** (0.088)	0.925*** (0.079)	0.836*** (0.074)	0.873*** (0.096)	0.865*** (0.081)
Observations	3344	3216	3296	3392	3376	3344
R <sup>2</sup>	0.081	0.038	0.054	0.030	0.126	0.023
Controls	✓	✓	✓	✓		

Table 10: Regression-DD estimates for reform effects with controls

Notes: The reference category for Models (1)-(4) is the control condition. The reference category for Model (5) is *Flat25,Prog*, and for Model (6) it is *Prog,Flat50*. Robust standard errors clustered on the subject level are in parentheses.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05.



	Diff-in-Diff – Labor Supply							
	$\leq 20$	$> 20$	$\leq 20$	$> 20$	$\leq 20$	$> 20$	$\leq 20$	$> 20$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post Reform	0.005 (0.005)	-0.049** (0.015)	0.005 (0.005)	-0.049** (0.015)	0.005 (0.005)	-0.049** (0.015)	0.005 (0.005)	-0.049** (0.015)
Prog,Flat25	-0.036 (0.021)	-0.007 (0.007)						
Prog,Flat50			-0.063** (0.023)	-0.013 (0.007)				
Flat25,Prog					-0.096** (0.032)	0.001 (0.006)		
Flat50,Prog							-0.130*** (0.037)	-0.004 (0.004)
Post $\times$ Treat	0.083*** (0.009)	0.046** (0.015)	0.028 (0.018)	0.011 (0.023)	0.039* (0.017)	-0.062*** (0.018)	0.119*** (0.025)	-0.062*** (0.017)
Age	-0.002 (0.001)	0.000 (0.000)	-0.000 (0.002)	0.001* (0.000)	-0.003 (0.002)	0.000 (0.000)	-0.001 (0.001)	0.000 (0.000)
Female	-0.002 (0.023)	-0.016 (0.009)	-0.010 (0.031)	-0.021 (0.014)	0.001 (0.029)	-0.007 (0.008)	0.015 (0.027)	-0.007 (0.008)
At Least Bachelor's Degree	0.024 (0.023)	-0.026* (0.013)	0.032 (0.031)	-0.018 (0.014)	0.050 (0.032)	0.009 (0.010)	-0.018 (0.030)	-0.012 (0.011)
Full-time Employed	0.009 (0.024)	0.018 (0.012)	-0.027 (0.031)	-0.014 (0.013)	-0.026 (0.034)	0.007 (0.009)	0.049 (0.031)	0.002 (0.009)
Below Median Income	0.007 (0.025)	-0.011 (0.010)	-0.024 (0.028)	-0.013 (0.014)	0.065 (0.036)	0.014 (0.010)	0.022 (0.032)	-0.002 (0.009)
MTurk Experience	-0.000 (0.001)	-0.000 (0.000)	0.001 (0.001)	-0.000 (0.000)	-0.000 (0.001)	0.000 (0.000)	-0.001 (0.001)	-0.001 (0.000)
Regime Fair	-0.013 (0.020)	0.010 (0.011)	0.011 (0.028)	0.000 (0.012)	0.021 (0.028)	0.015 (0.008)	0.014 (0.028)	0.005 (0.008)
Task Easy	-0.036 (0.025)	-0.010 (0.008)	-0.019 (0.029)	0.008 (0.017)	0.006 (0.033)	-0.013 (0.008)	-0.009 (0.033)	-0.004 (0.008)
Tax Understood	0.094 (0.054)	-0.008 (0.010)	0.013 (0.041)	-0.005 (0.014)	0.041 (0.051)	-0.012 (0.010)	0.064 (0.047)	-0.008 (0.010)
Financially Literate	0.048 (0.028)	0.010 (0.010)	-0.003 (0.032)	0.010 (0.013)	0.045 (0.036)	-0.000 (0.009)	0.008 (0.031)	-0.005 (0.009)
Risk Averse	-0.043 (0.022)	-0.001 (0.009)	0.030 (0.028)	-0.008 (0.012)	-0.010 (0.031)	0.001 (0.009)	0.005 (0.027)	0.004 (0.008)
Loss Averse	0.015 (0.030)	0.009 (0.012)	0.051 (0.036)	0.000 (0.015)	-0.017 (0.031)	0.008 (0.010)	0.024 (0.029)	0.010 (0.011)
Tax Honest	0.096 (0.055)	-0.000 (0.010)	-0.031 (0.043)	-0.008 (0.014)	-0.032 (0.037)	0.008 (0.012)	0.006 (0.035)	0.003 (0.010)
Conservative	-0.001 (0.023)	-0.006 (0.011)	-0.029 (0.028)	-0.006 (0.012)	-0.008 (0.028)	0.007 (0.008)	-0.013 (0.030)	-0.005 (0.008)
Constant	0.785*** (0.108)	1.003*** (0.022)	0.861*** (0.106)	1.017*** (0.030)	0.919*** (0.101)	0.955*** (0.025)	0.794*** (0.101)	1.001*** (0.025)
Observations	2786	558	2592	624	2156	1140	2224	1168
R <sup>2</sup>	0.116	0.132	0.050	0.096	0.089	0.224	0.083	0.249

Table 11: Regression-DD estimates for reform effect by reform affectedness with controls

Notes: The reference category is the control group *Prog,Prog*. For each treatment, we estimate two models in which we restrict the sample to labor supply decisions of either  $\leq 20$  or  $> 20$  labor units in  $T_0$ . Robust standard errors clustered on the subject level are in parentheses.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

	Extensive margin (ME) – Labor supply											
	No change				Reduce				Increase			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Prog,Flat25	-0.361*** (0.024)				-0.034 (0.019)				0.395*** (0.022)			
Prog,Flat50		-0.364*** (0.025)				0.080** (0.031)				0.310*** (0.026)		
Flat25,Prog			-0.309*** (0.031)				0.264*** (0.028)				0.056* (0.025)	
Flat50,Prog				-0.343*** (0.029)				0.258*** (0.029)				0.111*** (0.030)
Observations	1672	1608	1648	1696	1672	1608	1648	1696	1672	1608	1648	1696
Pseudo $R^2$	0.148	0.164	0.119	0.130	0.043	0.064	0.131	0.112	0.225	0.164	0.042	0.061
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 12: Probit estimates for extensive margin reform responses

Notes: The reference category for all Models is *Prog,Prog*. Robust standard errors clustered on the subject level are in parentheses.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05.

	Intensive margin – Labor supply											
	No change				Reduce				Increase			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Prog,Flat25	-0.004 (0.033)				-0.108 (0.099)				0.078** (0.026)			
Prog,Flat50		0.018 (0.025)				-0.234*** (0.063)				0.086*** (0.025)		
Flat25,Prog			-0.008 (0.027)				0.058 (0.059)				-0.050 (0.036)	
Flat50,Prog				0.019 (0.024)				0.091 (0.049)				-0.046 (0.041)
Observations	1121	1082	1160	1159	89	187	344	345	462	339	144	192
R <sup>2</sup>	0.014	0.023	0.047	0.022	0.194	0.267	0.080	0.051	0.086	0.089	0.069	0.168
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 13: OLS estimates for intensive margin reform responses

Notes: The reference category for all Models is *Prog,Prog*. Robust standard errors clustered on the subject level are in parentheses.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05.

	Diff-in-diff – Labor Supply			
	(1)	(2)	(3)	(4)
Post Reform	0.049*** (0.014)	0.049*** (0.014)	0.049*** (0.014)	0.049*** (0.014)
Prog,Flat25	0.066 (0.037)			
Prog,Flat50		-0.028 (0.038)		
Flat25,Prog			-0.320*** (0.030)	
Flat50,Prog				-0.331*** (0.028)
Post $\times$ Treat	-0.452*** (0.030)	-0.329*** (0.031)	0.218*** (0.031)	0.245*** (0.030)
Constant	0.436*** (0.072)	0.372*** (0.078)	0.536*** (0.070)	0.430*** (0.076)
Observations	3344	3216	3296	3392
R <sup>2</sup>	0.134	0.105	0.118	0.120
Controls	✓	✓	✓	✓

Table 14: Regression-DD estimates for bunching responses

Notes: The reference category for all Models is *Prog,Prog*. Robust standard errors clustered on the subject level are in parentheses.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05.

## B Instructions and screens

### Short survey on individual decision-making(~ 20 minutes)

**Description:** You will answer two brief questionnaires and make decisions. You will receive a bonus up to \$ 4.50 depending on your decisions in addition to a flat payment of \$1 for completing the study. Participation should take about 20 minutes. Please note that you can only click on the link once and that the Internet Explorer/ Microsoft Edge is not supported.

#### Instructions

Go to [Link](#) and follow the study instructions. Note the secret key found at the end of the study which you will need to complete the HIT.

\* 1. Enter the **SECRET KEY** (not your Worker ID) found at the end of the linked survey. Do not add any comment or text here


Submit

Powered by CloudResearch.com

Figure 9: Description of the human intelligence task (HIT) with the link to the experiment

Before you proceed to the survey, please complete the captcha below.

☐ I'm not a robot

  
reCAPTCHA  
[Privacy](#) - [Terms](#)

Please indicate your Mturk ID, so that we can pay you *after completion of this study*.

Continue

Figure 10: Welcome screen with a captcha

# Welcome

You will now take part in an academic research study from the University of Freiburg, Germany. With your participation you will make an important contribution to research. The study is about **individuals' behavior towards taxation**. You will participate in a demographic survey (Part I), a decision-making experiment (Part II), and a follow-up survey (Part III). Participation should take about **20 minutes**.

It is very important that you **read the instructions carefully** and that you complete the **entire study**, once you have started. This study will contain attention checks to ensure that you fully read the instructions. If you do not pass the attention checks, you will be excluded from the experiment.

For completing the study, you will receive **\$ 1**. Additionally, you can earn up to **\$ 4.50** depending on your decisions. Only after you completed the whole study, you will receive the MTurk code in order to receive your payment. Additional payments will be distributed as a bonus payment within three days upon completion of the study.

Confidentiality: You will remain **anonymous** throughout the study. Results may include summary data, but you will never be identified. All data collected are for **research purposes** only. By continuing, you consent to the publication of study results.

Participants' rights: **Your participation is voluntary**. You may stop participating at any time by closing the browser window or the program to withdraw from the study. If you stop participating in this study before completing it and you do not want your data to be used, please kindly contact us at [taxresearch@eucken.de](mailto:taxresearch@eucken.de).

Please also contact us at [taxresearch@eucken.de](mailto:taxresearch@eucken.de) if you have any questions.

By continuing, you indicate that you are **at least 18 years old**, have **read and understood the above**, and you voluntarily **agree to participate** in this study.

Continue

Figure 11: Introduction with participant's rights and consent

## Survey (Part I)

Please answer the following questions about yourself.

**Please indicate your gender**

**Please indicate your age**

**What is your highest educational degree?**

**What is your current employment status?**

Figure 12: Survey Ia



How high was your total household income, before taxes, last year (2021)?

How many hours per week do you spend working online on MTurk or similar platforms?

Now, please imagine the following situation. You can choose between different payoffs which are related to different probabilities of winning. Option A earns a payoff of \$1.90 with a probability of 20%. Option B earns a payoff of \$1.80 with a probability of 30%. Option C earns a payoff of \$1.70 with a probability of 40% etc. We are interested in whether you take the time to read the instructions; if not, some results may not tell us much about decision-making in the real world. To show that you have read the instructions, please check the third option below.

Please choose a combination of payoff and probability:

Option A: Probability 20% and \$1.90
Option B: Probability 30% and \$1.80
Option C: Probability 40% and \$1.70
Option D: Probability 50% and \$1.60
Option E: Probability 60% and \$1.50
Option F: Probability 70% and \$1.40

Figure 13: Survey Ib

## Instructions to the Decision-Making Experiment (Part II)

You will now participate in a decision-making experiment on behavior towards taxation. This experiment has a decision and a working stage:

The **decision stage** consists of 16 rounds. In each round you will **choose how much income you want to earn**. The income determines the number of tasks you have to complete later. The task is to transcribe words.

You will have to pay taxes on your income. **The tax rate may, but does not have to, vary from round to round**. Each of the 16 rounds is independent of each other.

In each round, you are first informed of the tax rate in this round and the income that you can earn. The income can be up to 600 cents. The higher the income you choose, the more tasks you will have to complete. The lower the income, the earlier you can finish the experiment. 20 cents correspond to 1 task.

After you have entered an income, the number of tasks and the due tax payment will be automatically calculated and shown on the screen. The tax payment equals the chosen income multiplied by the tax rate. After each round, you will receive information about your payoff. Your payoff is the chosen income minus the tax payment.

In the **working stage**, you will have to complete the number of tasks to earn the income that you indicated in one of the previous 16 rounds. This round will be randomly selected. It also determines how much your additional earnings from the experiment will be.

On the following screen, we will explain the working stage in more detail.

Continue

Figure 14: Instructions

## Working Stage

After the 16 rounds in the decision stage, you will have to work on the income you chose in one randomly selected round.

Your task is to transcribe text sequences. Each text sequence consists of 10 letters, see the example below. The number of tasks that you will work on depends on your decisions in the 16 rounds and on chance. A sequence is counted when you correctly typed in every letter.

**Please transcribe one text sequence below to familiarize yourself with the task.**

A "Continue" button will appear once you have correctly typed in one sequence.

**Correct Text Sequences: 0**

Text Sequence:

acyrgxrcqm

Answer:

Submit

Figure 15: Example task

# Control Questions

Please answer the following questions to verify that you understood the instructions. You have two attempts. Afterwards, you will be excluded from the experiment.

You may also review the instructions below.

---

## Instructions for Decision Stages

- Income up to 600 cents varies between the 16 rounds
- Tax rates may change between rounds, but do not have to
- You will choose an income, which determines the number of tasks you work on later
- 20 cents correspond to 1 task

Each round ends with information about what your payoff from that round would be.

## Instructions for Working Stage

You will have to complete the number of tasks to earn the income that you indicated in one of the 16 rounds. The round will be randomly selected. The chosen income in that round determines how much your additional earnings from this experiment will be.

---

Figure 16: Control questions

**Suppose you choose an income of 600 cents. How many text sequences do you need to complete in the working stage?**

1

30

60

100

**What does the additional payment for this decision-making experiment depend on?**

It depends on the income I choose in each round.

It depends on the tax rate.

It depends on the round that is selected by chance.

All of the above.

**Continue**

Attempts left to answer the control questions: 2

Figure 17: Control questions

## Round 1 of 16

**In this round, the tax rate is 25% for incomes equal to or below 400 cents. The tax rate is 50% on the entire income if the income exceeds 400 cents. For example, for an income of 420 cents, your tax payment would be 210 cents.**

You can earn an income of **480 cents**.

Please indicate whether you want to work for 480 cents or another income:

Income (in cents):

Number of text sequences to earn this chosen income:

Tax payment (in cents):

Continue

Figure 18: Decision-making stage with treatment sentence (example for progressive tax regime)

## Summary of Round 1 of 16

Chosen Income: 400 cents

Tax Payment: 100 cents

**If this round is chosen, your payoff equals 300 cents and you have to work on 20 text sequences.**

Continue

Figure 19: Summary after the decision-making stage

## Working Stage

Round 2 was randomly chosen to determine how much you need to work.

In round 2, you chose an income of 0 cents. To earn this income, you need to correctly transcribe 0 text sequences. When you reached the required number of text sequences, the "Continue" button will appear.

**Correct Text Sequences: 0**

Text Sequence:

nsmjqmgeip

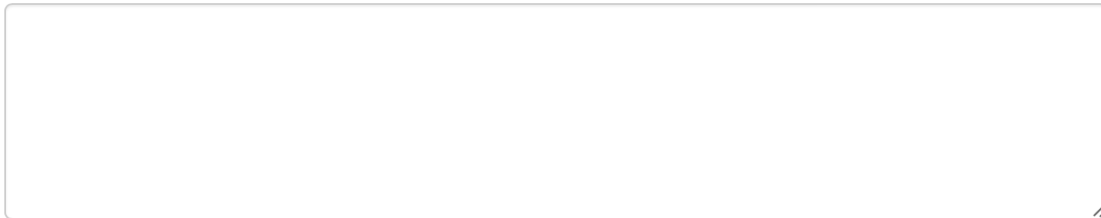
Continue

Figure 20: Working stage

## Survey (Part III)

In this third and last part of this HIT, we would like to know more about your decisions and attitudes.

**Please describe how your income was taxed in the different rounds in your own words.**



**If you think about choosing an income in the experiment, would you say that the tax payment was fair or unfair?**

Please indicate this on a scale from 1 to 5, where 1 means "not fair at all" and 5 means "very fair".

1 = Not fair at all  5 = Very fair

**How did you feel about the transcription task that you worked on?**

Please indicate this on a scale from 1 to 5, where 1 means "very tedious" and 5 means "not tedious at all".

1 = Very tedious  5 = Not tedious at all

Figure 21: Survey IIa



**Suppose you chose to earn an income of 420 cents. The tax rate is 25% for incomes equal to or below 400 cents. The tax rate is 50% on the entire income if the income exceeds 400 cents.**

**What would be the tax payment?**

105 cents
110 cents
210 cents

**Suppose you could earn 40 cents per task. The tax rate is 50% for all incomes. You can choose up to 30 tasks.**

Please indicate the number of tasks you would work on by moving the slider below.

-

tasks

+

Figure 22: Survey IIb

**For the next question, please imagine you are offered two options:**

Option 1: Have \$1 added to your compensation for this study.

Option 2: Have \$2 donated to a charity of your choice.

**Which of these options would you choose?**

Option 1

Option 2

---

**In the following table you find a list of coin tosses with different payoffs. The payoffs differ in how much you lose if the coin turns up heads. For each row, please indicate whether you would toss the coin or not.**

1. If the coin turns up heads, then you lose \$2; if the coin turns up tails, you win \$6.

Accept

Reject

2. If the coin turns up heads, then you lose \$3; if the coin turns up tails, you win \$6.

Accept

Reject

Figure 23: Survey IIc

3. If the coin turns up heads, then you lose \$4; if the coin turns up tails, you win \$6.

Accept

Reject

4. If the coin turns up heads, then you lose \$5; if the coin turns up tails, you win \$6.

Accept

Reject

5. If the coin turns up heads, then you lose \$6; if the coin turns up tails, you win \$6.

Accept

Reject

6. If the coin turns up heads, then you lose \$7; if the coin turns up tails, you win \$6.

Accept

Reject

Figure 24: Survey IId

**Suppose you had \$100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would have in the account if you left the money to grow?**

More than \$102.

Exactly \$102.

Less than \$102.

Do not know.

**Imagine that the interest rate on your savings account was 1 percent per year and inflation was 2 percent per year. After 1 year, would you be able to buy more than, exactly the same as, or less than today with the money in this account?**

More than today.

Exactly the same as today.

Less than today.

Do not know.

Figure 25: Survey IIe

**Do you think that the following statement is true or false? "Buying a single company stock usually provides a safer return than a stock mutual fund."**

True.

False.

Do not know.

---

**Have you ever submitted a tax declaration?**

Yes

No

**Please indicate whether you think it can always be justified, never be justified, or something in between on a scale from 1 (= never justifiable) to 10 (= always justifiable) to cheat on taxes if you have a chance.**

1 = Never justifiable



10 = Always justifiable

Figure 26: Survey II

**Please indicate, in general, how willing or unwilling you are to take risks.**

Please use a scale from 1 to 10, where 1 means you are "completely unwilling to take risks" and 10 means you are "very willing to take risks".

1 = Completely unwilling to take risks  10 = Very willing to take risks

**In political matters, people talk of "the left" and "the right." How would you place your views on this scale, generally speaking?**

Please use a scale from 1 to 10, where 1 means "left" and 10 means "right".

1 = Left  10 = Right

**As a result of the Covid-19 pandemic, did you...?**

☐ Lose all income from your job or business

☐ Earn some income but less than usual

☐ Earn about the same income as usual

☐ Earn more income than usual

**Do you have any comments concerning this study?**

Figure 27: Survey IIg

# Thank you for your participation!

For your payoff, round 2 was randomly chosen!

Your payoff in round 2 is 0 cents.

In total, you receive  $\$1 + \$0.00 = \$1.00$ .

**Please note this number for your payoff:**

**1000105**

Thank you for your participation!

Figure 28: Final stage with payoffs