# Behavioral Effects of Withholding Taxes on Labor Supply*

by

Johannes Becker†, Jonas Fooken‡ and Melanie Steinhoff §

**Abstract**

Income tax collection in most advanced economies uses third-party reporting and withholding at the employer level before the employee receives her wage income. Since withholding taxes do not necessarily reflect the true effective tax burden, they may give false signals on the net-of-tax pay. We report results of laboratory experiments in which labor supply effects of such misperceptions are tested. Withholding taxes (and resulting tax refunds) should be behaviorally neutral in the experiment, but our results suggest that withholding taxes reduce effort and tax adjustments lead to adjustments of effort for our most relevant group of participants, those motivated by monetary incentives. This indicates that withholding taxes may be behaviorally relevant and deserve attention of policy-makers.

**JEL classification:** H25, M41, G32

**Keywords:** Withholding taxes, experiment, tax perceptions

*We thank Benedikt Herrmann, Lars Siemers, Kay Blaufus, Denvil Duncan, conference participants at ACE Brisbane, ESA Sydney, ZEW Mannheim, IMEBESS Rome and seminar participants in Hamburg, Mannheim, Middlesex and Siegen for useful comments.

†Corresponding author, Institute of Public Economics, University of Münster, Wilmergasse 6-8, 48143 Münster, Germany, johannes.becker@wiwi.uni-muenster.de.

‡ Centre for the Business and Economics of Health, University of Queensland, jonas.fooken@gmail.com

§ Institute of Public Economics, University of Münster, Germany, melanie.steinhoff@wiwi.uni-muenster.de.

> Are you expecting a tax refund this season? If so, you're probably dreaming about all the possibilities of spending what feels like newly "found" money. However, it's important to remember that this is your money, that you earned – not a bonus. A tax refund is issued when you overpaid the government.
>
> Jon Lal, US News[2]

# 1 Introduction

Most wage earners receive an income tax refund at the end of the tax year. As the above quote suggests, this money is often treated as 'newly found', as a surprising one-time increase in income and evidence shows increasing consumption once the tax refund is paid (Shapiro and Slemrod, 1995). Since the refund can, in principle, be anticipated and does not represent an unpredictable increase in income, this finding is hard to reconcile with fully rational behavior (in the absence of liquidity constraints). Furthermore, if individuals change their behavior when the money is returned, it may well be that a similar change in behavior takes place when the money is taken away, i.e. when withholding taxes reduce income throughout the year. In particular, withholding taxes and their resulting tax adjustments at year end may affect labor supply by affecting the perceived immediate return from working (i.e. the perceived net wage).

In this paper, we investigate labor supply responses to withholding and

[2] `http://money.usnews.com/money/blogs/my-money/articles/2016-04-07/heres-how-you-should-spend-your-tax-refund` accessed April 7, 2016, at 10:43 a.m.

tax adjustments using a laboratory experiment, which allows us to make an experimental policy change that cannot easily be done in reality. Participants perform a real effort task and are paid a piece rate for each completed task while their income is taxed. Our treatments use different levels of withholding rates. In the absence of withholding, the wage earner receives 'monthly' gross wage payments; at the end of the 'year', she pays her tax liability as one large installment. With withholding, taxes are directly subtracted from the gross wage in each 'month'; at the end of the 'year', a tax adjustment takes place which implies an additional payment or a tax refund, depending on the level of the withholding rate and on effort. Thus, withholding taxes shift tax payments forward in time and increase their frequency. We use our experiment to test if withholding rates and tax adjustments influence labor supply decisions.

The experiment is designed such that withholding taxes and tax adjustments should be irrelevant for behavior. However, results of our experiment show that tax adjustments (refunds or additional payments) change effort provision in the subsequent period(s). After observing their true tax liability and end-of-year adjustments of their income, participants strongly adjust their labor supply. We interpret this as an update of their beliefs on the actual tax burden. Furthermore, on closer examination, we find that these findings are driven by participants who describe themselves as money-motivated (about half of our sample). Within this group, higher withholding rates significantly reduce effort in the subsequent period, while refunds increase effort. Because these two effects are countervailing, there is no effect of withholding on total effort (neither in the money-motivated subgroup nor

in the full sample). Overall, our results support the notion that incorrect perceptions of taxes due to withholding can influence labor supply decisions, making our results comparable to those by Abeler and Jäger (2015) whose paper is closest to ours. The intriguing point about withholding taxes is that they can be changed at little cost in terms of tax revenue, making them a simple tool for tax administrations or governments to influence economic outcomes. Policy-makers seem to have an intuitive understanding of this: Lacking support from Congress to lower the income tax rate, US President George H.W. Bush decided to lower the withholding rate to stimulate the economy.[3] Put apart caveats of translating lab results to real-world situations, our experiment allows for testing whether withholding taxes can have an effect on effort and informs policy-makers by showing potential labor market consequences of withholding rate changes.

The effect of withholding taxes on labor supply has, to the best of our knowledge, not been studied in the academic literature. However, our study relates to different strands of literature that we discuss in the next Section. Section 3 develops our predictions. Section 4 describes the experimental design and Section 5 the results. Section 6 discusses policy implications of our results and Section 7 concludes.

[3] In his State of the Union Address in January 1992, he said: "*This initiative could return about $25 billion back into our economy over the next 12 months, money people can use to help pay for clothing, college, or to get a new car.*"

## 2 Literature and hypotheses

Withholding taxes are a common feature of labor income taxation. Taxes on wage and salary income are levied in two steps. Firstly, the employer pays a withholding tax on the employee's monthly (or weekly/fortnightly) wage income. Secondly, once a year, the household submits a tax statement to determine the final effective tax rate and receives a refund or settles an additional liability. Thus, part of the tax is collected by the employer and the remaining balance is settled between the household and the tax administration.

### 2.1 Related literature

Economic theory suggests that the relevant price for the household's labor supply decision is the net wage rate. If true effective marginal tax rates are known, withholding rates should not matter for labor supply decisions as long as households are not liquidity constrained (beside a small interest rate effect, see Shapiro and Slemrod (1995)). What sounds straightforward in theory, becomes hard to grasp in real-world situations. In most countries the tax code is tremendously complex and true effective net wages are hard to calculate.[4] Complexity increases the cost of being fully informed. Individuals may therefore ignore some complexity of the tax system even if this implies suboptimal decisions, choosing to be rationally inattentive (Sims, 2003; Reis, 2006; Cheremukhin et al., 2011). They may also not know or misunderstand

[4] A tax system solving trade-offs between equity and efficiency (Mirrlees, 1971) is necessarily complex (Kleven and Kopczuk, 2011).

the true incentive environment (Chetty and Saez, 2013; Chetty et al., 2013).[5]

In many cases the net-of-withholding-tax wage indicated on the wage earner's monthly (or more frequent) wage sheet is a 'best guess' of the true effective tax rate. Households may, mistakenly or as a rule of thumb, interpret the net-of-withholding-tax wage as their true net wage. Such behavior would be in line with research showing that individuals tend to stick to salient features of a tax system (Chetty et al., 2009; Finkelstein, 2009) and may miss incentives if they are not salient (Chetty and Saez, 2013). However, if withholding taxes are levied at a constant rate, tax rates are progressive and income is variable, a withholding tax is essentially always imprecise. This imprecision is further increased if the household has multiple sources of income or if it consists of more than one person. It is this imprecision that makes withholding taxes behaviorally relevant - provided that households regard withholding as informative and do not recognize their true effective tax rates.

Our focus on withholding rate effects on labor supply links our study to two strands of literature. The first investigates the effects of tax withholding on household decisions. It shows that a reduction in the withholding rate influences consumption behavior by stimulating spending (Feldman, 2010). Moreover, withholding may increase tax compliance by limiting tax evasion opportunities through third-party reporting (Kleven et al., 2011; Slemrod

[5] Abeler and Jäger (2015) show that tax incentives are more likely ignored in complex environments. Furthermore, individuals' cognitive resources may be insufficient to prevent operational misunderstandings: Chetty et al. (2009) show that consumers have higher demand for goods when prices are presented exclusive of sales tax, even if they know that the sales tax applies and at which rate. Similarly, Feldman and Ruffle (2013) show that people buy more with tax exclusive prices than with tax inclusive prices.

et al., 2014), but potentially also by more voluntary compliance (Engström et al., 2015).[6] It may also affect the political salience of income taxation.[7] Cabral and Hoxby (2012) argue that the absence of withholding makes the property tax one of the most hated tax instruments, as in some US States it is levied all at once and perceived more strongly by taxpayers. These results indicate that high withholding rates may be desirable, at least for compliance and a reduced political cost of taxation.

The second strand analyzes how taxes affect labor supply decisions when the way of communicating them is varied. In particular, tax complexity may change how individuals respond to labor market incentives. Abeler and Jäger (2015), with a focus on experimental labor supply decisions close to our paper, show that under constant effective tax rates higher complexity leads to a misunderstanding of incentives and changes effort provision in an experimental real effort task. Similarly, Avram (2015) shows that framing a tax in different ways changes experimental effort.[8]

[6] Rees-Jones (2014) investigates 'overwithholding' by households and finds that households increase their tax avoidance if tax adjustments result in additional payments. Jones (2012) analyzes why households 'overwithhold' and argues that it is driven by inertia.

[7] Gamage and Shanske (2011) suggest to differentiate between economic salience depending on whether perceived taxes influence economic or voting decisions.

[8] Other studies show that decision-makers may become subject to a 'net wage illusion' by underestimating the impact of the tax rate, changing their labor supply if the gross wage differs, even though the net wage remains constant (Fochmann et al., 2013; Fochmann and Weimann, 2013) or when the wage is presented as a gross wage minus a tax instead of an equivalent non-taxed wage (Hayashi et al., 2012). These papers discuss explanations for such behavior, including anchoring on the base price, underestimating surcharges, preferences for presentation (dislike of 'taxes' which are labeled as such) and complexity aversion. Decision makers may also be more aware of a tax if it is connected to their current decision: Blumkin et al. (2012) show that participants reduce labor supply more when their labor income is taxed compared to when their consumption is taxed by the same amount. All these studies indicate that presenting income taxes as more immediate would lead to stronger labor supply responses.

## 2.2 Hypotheses

The above discussed literature suggests that salience matters when households have a prior belief on the effective tax rate, but are uncertain about the true final rate.[9] They use their wage and tax experience to update beliefs on the tax system. We adopt the approach by Feldman and Ruffle (2013) on household decisions to model these beliefs. Beliefs about taxes are updated based on income received from working and when the final tax liability is determined at the end of each year when the household receives a refund or pays an additional amount. Although the final tax liability may vary with the marginal tax rate being constant, an increase in the tax liability is interpreted as an increase in the marginal tax rate, and vice versa. The two settings (with and without withholding) give rise to different patterns of updating and, thus, to different behavioral patterns over different periods. In the setting with withholding tax, monthly income seems to be lower. At the end of the year, there is a tax adjustment, which makes the household adjust beliefs of the tax system. In our context, with differing withholding rates, we are able to experimentally test this model. Also see the online appendix (OA) for the description using a formal theoretical model.

First, if households are fully informed, neither the withholding tax nor the tax adjustment at the end of the year change the household's behavior.

Second, if households do not fully understand the tax system, but generally understand that the withholding tax is irrelevant for the calculation

[9] Neither the theory of rational inattention nor alternative approaches (Eliaz and Spiegler, 2011; Gabaix, 2011), make clear predictions how an otherwise irrelevant aspect becomes behaviorally relevant (see also Kőszegi and Szeidl, 2013; Bordalo et al., 2010).

of her effective marginal tax rate, beliefs and effort are not affected by the withholding rate. However, households may update beliefs and adjust effort after the final tax liability is announced and a tax adjustment is made.

Third, if households believe that the withholding tax reflects the effective tax rate, the withholding rate affects effort, as high withholding rates signal high effective tax rates. Additionally, effort is adapted after observing the tax adjustment as beliefs are updated again. These three cases can be formulated as a null hypothesis (the fully informed case) and two competing hypotheses.

$H_0$ Tax adjustments and the withholding tax do not affect behavior.

$H_{1a}$ Tax adjustments affect behavior but withholding rates do not.

$H_{1b}$ Both the withholding tax and tax adjustments affect behavior.

## 3 Experimental design

In the experiment, labor supply was measured using a real effort task that required participants to decode sequences. See Chow (1983) or Erkal et al. (2011) for the use of decoding tasks in real effort experiments.[10] 411 voluntary student participants (45% female, average age 22) from various disciplines were recruited using ORSEE (Greiner, 2015). The experiment was programmed in z-Tree (Fischbacher, 2007). Sessions were conducted between November 2014 and July 2015 at the University of Münster.

[10] We chose this task as we expected it to lead to a greater variance in effort provision compared to other tasks. Using decoding tasks may include an element of ability; however, as we are interested in differences between treatments, this is does not create undesired effects. We show later in the paper that low-ability participants are not driving our results.

Participants were asked to individually go through the instructions on the real effort task. They had to solve three control questions before continuing. Subsequently, participants were informed that they would be proceeding through two rounds of solving tasks with three periods each.[11] Each period of decoding lasted three minutes. Participants could solve as many tasks as they liked. The task required decoding sequences of eight Greek letters using a decoding table at the lower part of the screen assigning numbers to Greek letters (see OA for an example). Sequences and the decoding table changed after a sequence had been correctly decoded. For each correctly decoded sequence participants received a gross piece rate of either 50 or 30 Euro cents depending on the treatment.

After each round, a participant's total income from a round (of three periods) was subject to a progressive income tax. Table 1 shows the tax scheme as provided to participants. Additionally, a detailed example described how the final tax would be calculated. Tax rates were chosen to ensure sufficient variation in the effective average tax rate. We used a pretest to estimate expected performance and most participants (90% of participants in the first and 95% in the second round) decoded at least 17 sequences in one round, hence reaching the highest bracket (the average number was 26 per round).

In treatments with withholding, participants received additional information on how the withholding tax would be deducted from their earnings at the end of each period. The instructions explicitly explained that withholding tax payments were only provisional and that the final tax payment

[11] Each period may be interpreted as a month and three periods as a year, however, they were not referred to as months and years in the instructions.

Table 1: Marginal tax rates by bracket

| Solved tasks | Tax rate |
|---|---|
| $0-4$ | 0% |
| $5-10$ | 20% |
| $11-16$ | 36% |
| $17-\infty$ | 50% |

would be (re)calculated according to the progressive tax scheme at the end of the round, i.e. after three periods. It was also explained that this calculation might result in a refund or an additional payment. Since interest rate effects and other confounding factors, such as having income available for consumption or using high withholding as a savings commitment device, are absent in the experiment, the withholding rate has no impact on the final payment to participants.

We used several treatments, all between subjects. First, we varied the withholding rate, using rates of 0%, 20% and 50%. Second, we had treatments with and without a minimum effort option (see below). Third, we varied the payment per correctly solved task between 30 cents and 50 cents; this was done to contrast changes in *perceived* economic incentives to changes in *effective* ones. Table 2 summarizes the different treatment combinations and the number of participants and sessions per treatment. Participants were randomly assigned to a treatment depending on the session they attended.[12] Differences in the number of participants per treatment are mainly due to the fact that the number of participants coming to a session was not fully

[12]That is, when sessions were scheduled for a period of one or several weeks of data collection, treatments were assigned to sessions, while avoiding clustering treatments by certain characteristics, such as morning or afternoon, specific weekdays, etc.

predictable (and the number of participants per session was not fixed).[13]

Table 2: Number of participants [sessions] per experimental treatment

| Minimum Effort | Payment per task | Withholding rate | | | |
|---|---|---|---|---|---|
| | | 0% | 20% | 50% | Total |
| No | 30 cents | 37 [2] | 39 [2] | 44 [3] | 120 [7] |
| No | 50 cents | 35 [2] | 38 [2] | 32 [2] | 105 [6] |
| Yes | 30 cents | 79 [8] | | 107 [7] | 186 [15] |
| | Total | 151 [12] | 77 [4] | 183 [12] | 411 [28] |

In treatments with a minimum effort option, participants could opt out of the piece rate and choose a fixed payment for decoding two sequences (it was possible to decode additional sequences but these were not remunerated; in our data all participants who chose minimum effort solved at least two sequences). The decision on minimum effort had to be made before the respective period started. We thus added an extensive margin to the labor supply decision problem. The payment for the minimum effort option was exempt from tax. The level of payment, 120 Cents, was chosen such that it is an attractive option if effort is perceived to be very costly or the tax as high. If participants misinterpret the 50% withholding tax as final, they believe that they have to solve at least 9 sequences (remunerated at 30 cents each) to outperform the minimum effort option. However, 50% is only the marginal tax rate in the highest bracket, which most participants only reach at some later time during the third period of a round. The minimum effort option was available from period 2 onwards. This ensures that all participants observe

[13]The experiment was the first experimental economic study at this laboratory, therefore participant attendance habits were not known before the beginning of the data collection. Further sessions were added later in the year to reach at least 32 participants in each treatment arm and enough cases of minimum effort choices. However, also at later dates of data collection a range of different treatments were included.

the withholding rate at least once (i.e., after the first period). Furthermore, this way participants learn that it is easy to solve more than two sequences in one period (the average number of solved sequences was 8.7).

In treatments with the minimum effort option, we did not vary the payment per task because the previous results had not indicated strong effects of paying 30 cents instead of 50 cents. Furthermore, we only used withholding rates of 0% and 50% to increase the statistical power of comparisons. As only a small fraction of participants chose minimum effort at least once, larger samples are necessary to make sound comparisons. For the minimum effort option we increased the number of subjects until a sufficient number of participants had chosen minimum effort at least once. The number of participants required in the minimum effort treatment was estimated after running the first sessions, aiming at 30 individuals choosing minimum effort at least once.[14]

All participants were able to keep track of their gross income earned in a period (in real time) at the top of the decision screen.[15] The treatments differ in the time profile of information. After each period, the no-withholding group received a summary of gross income earned in a period, whereas the withholding group was provided with a summary of gross income minus the withholding rate and the withheld tax payment. Furthermore, participants in all treatments (with and without withholding) were reminded that these income and tax figures were preliminary and that the final tax liability would

[14] One more session was run for the high withholding rate by chance, as this session had already been scheduled, resulting in 28 participants more in this treatment arm.

[15] This information was provided independent of the withholding rate. In case minimum effort was chosen the number of correctly solved sequences as well as information that two sequences needed to be solved was provided.

be calculated at the end of the round. In case of the minimum effort option, the numbers of correctly solved tasks as well as the tax free payment of 1.20 Euro was displayed. After each round, the final tax liability for the three preceding periods was calculated and displayed. Withholding tax payments were credited against the final tax liability and tax adjustments were shown. This calculation results in either a tax refund or an additional tax payment. If minimum effort was chosen in some periods, income from these periods was added to the final income of the round. Due to this design responses to withholding taxes will always be observable in the subsequent period (in the first period effort should be the same across treatments). Thus, our identification strategy rests on the comparison of participants which have perfectly identical incentives to exert effort but are provided with different streams of information on income from effort.

Each experimental session lasted approximately one hour including instruction time and a post-experimental questionnaire. Participants received a show up fee of 5.00 Euro plus earnings from solving tasks during the experiment, earning on average 17.87 Euro. After all participants had finished, they were given a short questionnaire on personal background, motivation to participate in the experiment and attitude towards taxes (see OA for a list of questions).

Our pretest, intended to estimate the expected number of solved sequences per participant, indicated that numerous participants mainly tried to solve as many sequences as possible, independent of incentives. To identify this group of *intrinsically ambitious* participants as well as those who would be motivated to exert effort in response to economic incentives, we

included a question in our post-experimental questionnaire to identify these and other types of individuals (we added two further categories of potential motivation). Although rarely studied as the main variable of interest, the existence of types is widely documented in the experimental economic literature (e.g., Fischbacher and Gächter, 2010; Kurzban and Houser, 2005). Table 3 provides information on the sample of participants based on their motivation to solve tasks. While there are more individuals motivated by money and less ambitious participants in the 20% withholding treatment, no relationship between these two motivations and the withholding rate is observable when testing for statistical significance. Testing for whether treatments induce certain motivations does not show any significant bias.[16] See the OA for detailed results, showing that motivational type does not appear to be influenced by treatment (supporting our assumption that these self-assessed traits can be considered fixed). Similarly, the fraction of participants motivated by money is not significantly more common ($p>0.151$) in the 20% treatment. Further testing also shows that performance in the experiment is significantly higher for ambitious participants ($p\leq0.001$) who solve on average 6.36% more than the other groups.

In the analysis below, we differentiate in large parts of our analysis by these self-chosen types and focus on those motivated by money. Our conjecture is that those motivated by money react more strongly to economic incentives. Our results show that they do, being the driving factor behind

[16] Only the presence of the outside option is correlated with a higher number of participants reporting their main motivation to solve tasks as *fun*, which should not have any effect on the results, given the low number of participants in this category and the fact that results do not depend on pooling these participants with money-motivated or non-money-motivated subjects.

Table 3: Self-reported motivations of participants

| | All | 0% | 20% | 50% |
|---|---|---|---|---|
| Earn as much money as possible | 52% | 50% | 60% | 50% |
| Ambition | 41% | 42% | 32% | 44% |
| Fun | 5% | 7% | 4% | 4% |
| To do what the experimenter asked me to do | 2% | 1% | 4% | 2% |

any findings on the effects of tax adjustments and withholding, but also by having higher effort in sessions with a higher piece rate – in contrast to ambitious participants whose effort is rather lower when they are paid more for their effort.

# 4 Experimental results

## 4.1 Baseline results

We start by reporting the basic properties of our experimental data. There is an upward trend in effort over the six periods, which we interpret as learning or practice, whereas better performance in the previous period improves practise and positively influences the score in the following period(s). Average effort is 7.8 tasks solved (standard deviation = 1.8) in period 1 and 10.1 (s.d.=2.4) in period 6. The average numbers per round are 8.5 (s.d.=1.9) in round 1 and 9.7 (s.d.=2.1) in round 2.

The withholding rate has no signiqficant impact on aggregate effort (9.1 tasks per period for 0%, 8.9 for 20% and 9.2 for 50%, all pairwise comparisons between treatments using t-tests and ranksum tests with $p \geq 0.2$). Given that participants faced the same tax incentives, this may be considered unsurprising. However, it does not necessarily mean, though, that withhold-

ing taxes have no impact at all because withholding taxes affect the time profile of information received by participants. If net-of-withholding wages are perceived as an approximation of effective net wages, participants in the 50% treatment perceive their wages to be lower than those in the 20% or 0% treatment. Our theory implies that participants update their beliefs after having received income after each period and in period 4 when having received a refund or paid additional taxes. Therefore the question is whether new information leads to an adjustment in effort.

In period 1, each participant has the same set of information (recall that participants do not experience withholding taxes before the end of period 1) and, reassuringly, effort in this period is virtually the same across withholding treatments.[17] We subsequently loo at between-period changes in effort. [18] Taking this as a starting point, the treatments vary in the information reported to the participants. In periods 2 and 3 as well as 5 and 6, participants have just received their wage net of withholding tax or their gross wage, depending on the treatment. In period 4, participants have just experienced the tax adjustment, i.e. they had to pay additional taxes or got a refund. Hence, we have 'Withholding' periods (2,3,5,6) and a 'Tax adjustment' period (4).

[17] All $p \geq 0.44$ for any t-test of pairwise comparisons between treatments and also small in magnitude, with 2.5% deviation at the mean being the largest difference.

[18] The focus on changes instead of absolute levels is due to the nature of data generated in our real effort task and due to the conjectures effect of withholding. Firstly, effort provided in real effort tasks tends to be concentrated around maximum effort. Hence, treatment differences will mainly reflect how much participants try to extend their maximum score, which will be done by increasing their score from one period to the next (or reduce effort in case demotivating information has been received). Secondly, our design only changes the information flow between periods. Therefore, new information will only lead to an adaption of behavior compared to the previous period. Finally, due to the nature of withholding taxes, withholding in normal months and tax adjustments at year end have effects in the opposite direction. All these factors are best reflected when using between-period changes in effort as the variable of interest.

We can also use a more narrow definition of more pure 'Withholding' periods (1,3), as later periods may be influenced by having experienced tax adjustments before, and this experience may have a lingering-on effect.

If information is behaviorally relevant, we should expect individuals to adjust their effort differently in each treatment. Figure 1 presents the between-period changes in effort for periods 2 and 3, and period 4 using the more narrow definition. The corresponding figure with the wider definition is included in the appendix. Figure 1 shows changes for the full sample (left panel) and the money-motivated subsample (right panel), previewing our main findings, showing no visible effect of the withholding rate on effort for the full sample, and a strong difference between high and low withholding treatments and their resulting tax adjustments. The right panel refines this picture and shows effects of both the withholding rate and tax adjustments for the money-motivated subsample.

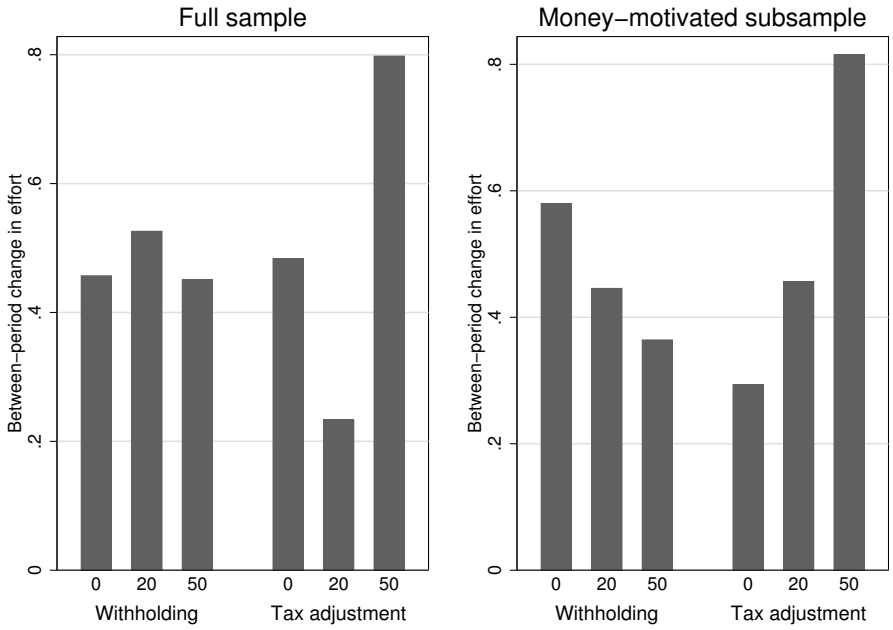
We interpret this as first evidence that the difference in information flows due to withholding rates may affect effort, at least via the tax adjustment. To further examine this effect, we turn to regression analysis. We continue to focus on between-period changes as the main outcome variable.[19] On average, participants increase their effort by 0.45 (standard deviation: 1.78) with a small (but significant) decreasing trend towards later periods.[20] We test our hypothesis in a regression model which can be described as

$$\Delta \text{ Effort}_t = \alpha + \text{X}\beta + \gamma \text{ tax adjustment} + \delta \text{ withholding rate} + \epsilon$$

[19] The OA shows that the main result can also be shown using effort levels as the outcome variable; that is, a significant effect of tax adjustment on effort for the full sample and a stronger effect for those who are money-motivated.

[20] The corresponding number for the subsample of money-motivated participants is a mean of 0.39 with a standard deviation of 1.80.

Figure 1: Average change in effort to the previous period




The figure shows the change in average effort (i.e. the change in the number of sequences solved) for the full sample (left panel) and the money-motivated subsamples (right panel). Within each sample, the average effort change of the periods 2 and 3 ("Withholding") and the average effort change in period 4 ("Tax adjustment") are reported.

, whereas $\gamma$ and $\delta$ are our main variables of interest and X$\beta$ includes a vector of control variables, depending on the specification. Table 4 shows the corresponding regression results for the full sample. The between-period change in effort is regressed on the tax adjustment (with negative values implying a refund), the withholding rate (0, 0.2 or 0.5) and other controls.[21]

[21]For the following regressions we pool treatments with and without the minimum effort option. Only very few participants chose minimum effort and appear to be those in the distribution of participants who would provide low effort also in sessions without this option. Among those who chose minimum effort, a large fraction also exerts more than minimum effort, again making effort of minimum-effort choosers very comparable to

Columns 1 and 2 report the results of the simplest regression using only the tax adjustment (as parametric variable) in period 4 as the main explanatory variable, as well as individual fixed (1) and random effects (2) linear models. In column 3, a lagged variable for the tax adjustment is added.[22] Participants adjust their effort after period 3 according to the information received at the end of the first round. A refund increases effort, an additional tax payment reduces effort. Column 4 adds the withholding rate (as parametric variable),[23] column 5 the pay level (50c or 30c) and column 6 potential interactions between the two.[24] The tax adjustment effect remains robust, whereas the effect of the withholding rate is insignificant (p=0.11 for specifications 4-5 and p=0.18 for specification 6).[25]

While the significant effect of tax adjustments supports our claim that withholding taxes are behaviorally relevant, the findings of Table 4 point to a conflicting result. If only tax adjustments trigger a behavioral response while the withholding rate does not have an effect during normal periods,

low-effort individuals in the distribution of sessions without the minimum effort option. The outside option hence serves more as a sorting device than changing the distributional structure of effort choices. The OA also shows that the main findings of tax adjustments and withholding for money-motivated participants is very similar (and differences statistically insignificant) between treatments with and without the minimum effort option. We pool the treatments with and without the minimum effort option to increase our sample size and statistical power.

[22]The OA includes more regression estimates with additional lagged variables for the tax adjustment and binary indicator variables for the withholding rates.

[23]Using the withholding rate as an indicator variable instead shows that any withholding effect would be driven by the high withholding rate. Using the continuous variable allows us to include a combined variable of the withholding rate and of paying 50 cents instead of 30 cents per task which can still be meaningfully interpreted. Including this combined variable is the theoretically most correct approach, as it informs about the displayed net amount paid per task which the participant observes each round.

[24]Since the withholding rate affects the *perceived* net wage differently if the gross wage changes, we control for differential effects by including an interaction variable.

[25]Using the tax refund per solved task instead of the raw value of the total tax refund provides the same result.

Table 4: Determinants of between-period changes in effort for the full sample

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Tax adjustment | -0.13*** | -0.10*** | -0.09** | -0.12** | -0.12** | -0.12** |
| | (0.05) | (0.04) | (0.04) | (0.05) | (0.05) | (0.05) |
| Tax adjustment$_{t-1}$ | | | -0.01 | | | |
| | | | (0.04) | | | |
| Withholding rate | | | | -0.22 | -0.22 | -0.19 |
| | | | | (0.14) | (0.14) | (0.14) |
| 50c per sequence | | | | | 0.00 | 0.04 |
| | | | | | (0.04) | (0.06) |
| Withholding rate | | | | | | -0.18 |
| × 50c per sequence | | | | | | (0.21) |
| Fixed/random effects | FE | RE | RE | RE | RE | RE |
| Constant | 0.46*** | 0.46*** | 0.36*** | 0.52*** | 0.51*** | 0.51*** |
| | (0.01) | (0.02) | (0.02) | (0.04) | (0.05) | (0.05) |
| n (individuals) | 411 | 411 | 411 | 411 | 411 | 411 |
| N (observations) | 2055 | 2055 | 1644 | 2055 | 2055 | 2055 |

Linear random effects regressions, dependent variable: Change (from preceding period) in sequences solved, standard errors (in brackets) clustered by individual. * indicates 10%, ** 5% and *** 1% significance. The variable 'Withholding rate' takes the values 0, 0.2 and 0.5; 'Tax adjustment' describes the (positive or negative) adjustment observed in period 4 (the variable is equal to zero in all other periods); and '50c per sequence' is a dummy variable equal to 1 if the payment was 50 cents instead of 30 cents.

total effort should be higher in treatments with higher withholding because tax adjustments are a direct function of withholding. However, this is not the case, as shown above. A potential explanation for this puzzle may be that the withholding rate has a negative effect on effort, but this effect is imprecisely measured in the full sample. In the following, we will therefore split the sample to reconcile these two mutually contradicting findings, reverting to our pre-experimental observation of some participants being seemingly unresponsive to any incentive, and try to identify individuals reacting to experimental incentives.

### 4.2 Subsamples by motivation

As our pre-test had indicated that some participants appeared to be unresponsive to economic incentives which can be understood as reflecting the heterogeneity in motivations to work (earn money, work ethic norms, self-fulfillment, etc.) in real-world labor markets. Heterogeneity of reactions in terms of performance to monetary incentives in real effort experiments is commonly observed in the literature (see Camerer and Hogarth (1999) for a review of heterogeneous effort responses to financial incentives. We therefore use our data on self-assessed motivation to solve tasks to identify participants who may be more responsive to incentives and may therefore drive our results. In the following, we differentiate between individuals who indicate to be mainly motivated by money and those who are motivated otherwise (i.e. by ambition, fun and to please the experimenter). Like on real-world labor markets, extrinsic motivation may affect individual behavior with differing strength and the heterogeneity in responsiveness to incentives has been shown in the literature.

Table 5 shows the regression results for the subsample of money-motivated participants. Not only is the effect of tax adjustment significant and stronger than in the full sample, but the withholding rate has a significant effect (specifications MM4-MM6). Additionally, the amount paid per sequence (MM6) have a meaningful and sizeable effect, which is not observable to the non-money-motivated sample. [26] This resolves the conflicting results in the

[26]The OA provides the equivalent results for the subsample of non-money-motivated participants. For them neither the tax adjustment nor the withholding rate have a significant effect on effort. In other words, the effects found in the full sample seem to be entirely driven by the participants that assess themselves as mostly money-motivated.

full sample: The withholding tax reduces effort, the tax refund increases effort. In total, withholding rates have no impact on total effort. However, both findings are only visible for those participants reporting to be money-motivated and appearing to respond to incentives.

Table 5: Determinants of between-period changes in effort of money-motivated participants

| | (MM1) | (MM2) | (MM3) | (MM4) | (MM5) | (MM6) |
|---|---|---|---|---|---|---|
| Tax adjustment | -0.18*** | -0.12*** | -0.11** | -0.16*** | -0.17*** | -0.17*** |
| | (0.06) | (0.05) | (0.05) | (0.06) | (0.06) | (0.06) |
| Tax adjustment$_{t-1}$ | | | -0.04 | | | |
| | | | (0.06) | | | |
| Withholding rate | | | | -0.44** | -0.43** | -0.36* |
| | | | | (0.20) | (0.20) | (0.20) |
| 50c per sequence | | | | | 0.10 | 0.21** |
| | | | | | (0.06) | (0.09) |
| Withholding rate | | | | | | -0.47 |
| × 50c per sequence | | | | | | (0.33) |
| Fixed/random effects | FE | RE | RE | RE | RE | RE |
| Constant | 0.41*** | 0.41*** | 0.33*** | 0.53*** | 0.50*** | 0.48*** |
| | (0.01) | (0.03) | (0.04) | (0.06) | (0.06) | (0.06) |
| n (individuals) | 213 | 213 | 213 | 213 | 213 | 213 |
| N (observations) | 1065 | 1065 | 852 | 1065 | 1065 | 1065 |

Linear random effects regressions, dependent variable: Change (from preceding period) in sequences solved, standard errors (in brackets) clustered by individual. * indicates 10%, ** 5% and *** 1% significance. The variable 'Withholding rate' takes the values 0, 0.2 and 0.5; 'Tax adjustment' describes the (positive or negative) adjustment observed in period 4 (the variable is equal to zero in all other periods); and '50c per sequence' is a dummy variable equal to 1 if the payment was 50 cents instead of 30 cents.

### 4.3 Robustness

We scrutinize the robustness of our findings on the effect of the withholding rate in several ways. We first investigate if the effect of adjustments lingers on and influences decisions in further periods. Results from this analysis show that such a continuation effect, although being in the right direction, is not consistently significant (MM3 in Table 5; also see OA for alterna-

tive specifications). We also test whether the effect of tax adjustments is significant both for refunds and additional tax payments, or if there is a breakpoint (e.g., due to loss aversion). We find that the effect is about of the same magnitude for refunds and additional payments as there is no significant interaction effect (see OA).

We subsequently introduce further control variables based on the experimental structure and our post-experimental questionnaire. These include the score of the first period ('Initial score') as an indicator of ability, period dummies to account for possible time trends and demographics (gender and age). Specification (MM7) in Table 6 shows the outcome of this procedure and represents a possibly 'richest' specification. As can be seen, the initial score and period controls are significant, while demographics are not. The main result regarding tax adjustments and withholding remain. This result does not change when including further control variables based on the post-experimental questionnaire (see OA).

We also test if participants' skill levels drive the effect of a reaction to withholding rates. To answer this question we split the sample of money-motivated participants once more into high and low performers using the score in the first period as an indicator of ability, see specifications (MM8) and (MM9) of Table 6. Using the time used for solving the test questions or self-reported math skills in the post-experimental questionnaire instead provides qualitatively the same results (see OA). As can be seen, rather high-ability individuals seem to drive the result: For them the effect is at least not weaker than for low-ability participants.

Finally, we check if learning about the withholding rate in the second

Table 6: Drivers of between-period changes in effort for money-motivated participants

| | | Initial score | | Round | |
|---|---|---|---|---|---|
| | | low | high | first | second |
| | (MM7) | (MM8) | (MM9) | (MM10) | (MM11) |
| Tax adjustment | -0.20*** | -0.14 | -0.22*** | | -0.24*** |
| | (0.07) | (0.10) | (0.08) | | (0.07) |
| Withholding rate | -0.39* | -0.29 | -0.52* | -0.34 | -0.51* |
| | (0.21) | (0.27) | (0.28) | (0.31) | (0.28) |
| 50c per sequence | 0.22** | 0.20 | 0.23** | 0.03 | 0.38*** |
| | (0.10) | (0.13) | (0.10) | (0.18) | (0.13) |
| Withholding rate | -0.50 | -0.20 | -0.75** | -0.18 | -0.78* |
| × 50c per sequence | (0.33) | (0.44) | (0.32) | (0.63) | (0.45) |
| Initial score | -0.04** | -0.07** | 0.05 | -0.23*** | 0.08*** |
| | (0.02) | (0.03) | (0.05) | (0.04) | (0.02) |
| Female | -0.02 | -0.02 | -0.01 | -0.04 | -0.01 |
| | (0.06) | (0.08) | (0.08) | (0.13) | (0.08) |
| Age | 0.01 | -0.00 | 0.02* | 0.04* | -0.01 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) |
| constant | 0.91*** | 1.83*** | -0.81** | 1.83*** | -0.22 |
| | (0.27) | (0.36) | (0.38) | (0.56) | (0.33) |
| Period Dummies | Yes | Yes | Yes | Yes | Yes |
| n (individuals) | 213 | 130 | 122 | 213 | 213 |
| N (observations) | 1065 | 650 | 610 | 426 | 639 |

Linear random effects regressions, dependent variable: change (from preceding period) in sequences solved; standard errors clustered by individual. * indicates 10%, ** 5% and *** 1% significance. The variable 'Tax adjustment' describes the (positive or negative) adjustment observed in period 4 (the variable is equal to zero in all other periods). Initial score describes the effort in the first Period; other control variables as indicated.

round (or 'year') leads to a disappearance of the effect. If this were true, the withholding rate effect would be only short-lived (and of little relevance in real world tax systems). However, again we find that the effect does not become smaller (specifications (MM10) and (MM11)). Hence, both higher ability and learning do not seem to correct the misperception of income and taxes due to withholding.

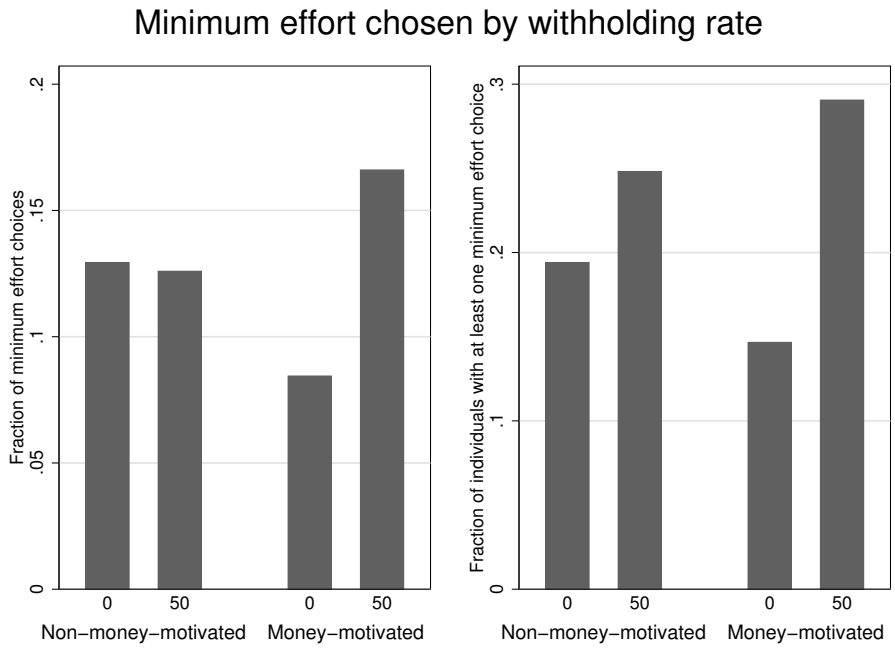
### 4.4 Minimum effort choices

Do high withholding rates cause participants to opt out of the pay-for-performance scheme and choose the minimum effort option? As shown in Table 2, 186 participants could chose minimum effort from Period 2 onwards. Concentrating on those participants mainly motivated by money, this sample is reduced to 98 participants. Figure 2 shows the frequency of minimum effort choices as well as the fraction of individuals who chose minimum effort at least once, separated by those mainly motivated by money and others.

While there is virtually no effect for non-money-motivated participants, the frequency of minimum effort choices when withholding is high (17%) is more than twice the frequency under no withholding (8%) for the money-motivated. A similar, but less striking result is visible when looking at the fraction of individuals with at least one minimum effort choice. This indicates that withholding influences the decision to opt out of the performance-based income scheme. However, analysis of the data needs to consider that these results are influenced by individuals who completely opt out of the piece rate, while others choose minimum effort only once or twice.

We first test the higher propensity to opt out of the piece rate using

Figure 2: Choice of the 'minimum effort' option




The figure shows the fraction of times in which participants chose the minimum effort when this option was available, for the subsamples of non-money-motivated and money-motivated participants and, within these two subsamples differentiated for the 0 per cent and 50 per cent withholding rate treatments. For example, money-motivated participants chose the minimum effort option in about 17 per cent of all periods with a 50 per cent withholding rate.

several tests. We first use t-tests and Wilcoxon ranksum test of participants choosing minimum effort at least once, both indicating a significant difference (t-test: p=0.047; ranksum: p=0.048) for the full sample, as well as for the subsample of money-motivated participants (t-test: p=0.036; ranksum: p=0.036). We then calculate the total number of minimum effort choices by individual and test these across treatments, showing an insignificant effect in a t-test (p=0.234) for the full sample and a marginally significant one in

a ranksum test (p=0.087), while being marginally significant in a t-test for the money-motivated subsample (p=0.095) and insignificant in a ranksum test (p=0.130). However these tests cannot fully reflect the repeated-choice structure of the data.

We therefore scrutinize the difference in a regression analysis, using linear and probit random-effects models. Results from this analysis for the money-motivated participants are shown in Table 7. They are, however, inconclusive, with the significance level being either marginally significant or insignificant depending on the specification. The corresponding results for the full sample (included in the OA) show no significant overall effect. Hence, even for money-motivated participants we cannot firmly conclude if withholding or tax adjustments lead to significantly more minimum effort or not and the effect is insignificant when including all participants.

## 5 Discussion: Implications and external validity

Our study shows that tax withholding can be behaviorally relevant by changing labor supply, most strongly through resulting tax adjustments at year end but potentially also throughout the year. Our results suggest that participants in our experiment update their beliefs on the tax burden by putting too much weight on currently experienced tax payments. Our hypothesis $H_0$ of withholding and adjustments being behaviorally neutral is rejected in favor of $H_{1a}$ for all participants, due to the significant effect of tax adjustments. As those participants who drive our rejection of $H_0$ also react significantly to withholding over the periods, $H_{1b}$ appears as the most appropriate alter-

Table 7: Determinants of minimum effort choices of money-motivated participants

| | (MM10) | (MM11) | (MM12) | (MM13) | (MM14) | (MM15) |
|---|---|---|---|---|---|---|
| | Linear model | | | Probit model | | |
| Withholding rate=50 | 0.08* | 0.08* | 0.10** | 0.59* | 0.60 | 0.75* |
| | (0.05) | (0.05) | (0.05) | (0.36) | (0.37) | (0.39) |
| Tax adjustment | | 0.00 | 0.01 | | 0.02 | -0.18 |
| | | (0.01) | (0.01) | | (0.15) | (0.33) |
| Initial score | | | -0.04*** | | | -0.32*** |
| | | | (0.01) | | | (0.11) |
| Female | | | -0.07 | | | -0.49 |
| | | | (0.05) | | | (0.38) |
| Age | | | 0.00 | | | 0.02 |
| | | | (0.01) | | | (0.07) |
| Period Dummies | NO | NO | YES | NO | NO | YES |
| Constant | 0.08*** | 0.08*** | 0.42* | -2.13*** | -2.13*** | 0.35 |
| | (0.03) | (0.03) | (0.23) | (0.33) | (0.34) | (1.66) |
| n (individuals) | 98 | 98 | 98 | 98 | 98 | 98 |
| N (observations) | 490 | 490 | 490 | 490 | 490 | 490 |

Linear and probit random effects regressions, dependent variable: Minimum effort option chosen (=1) or not(=0), standard errors (in brackets) clustered by individual in the linear model. * indicates 10%, ** 5% and *** 1% significance. The variable 'Withholding rate=50' is a binary variable equal to one when the withholding rate is 50%; 'Tax adjustment' describes the (positive or negative) adjustment observed in period 4 (the variable is equal to zero in all other periods); 'Initial score' describes the effort in the first period; other control vairables are as indicated.

native to $H_0$ for money-motivated participants.

## 5.1 Economic significance

While we show statistically significant effects of withholding on labor supply, we also consider the economic significance of our results. A withholding rate of 50 per cent reduces the average increase in effort by 0.18 (=0.36 * 0.5) sequences (see MM6), which is large relative to the corresponding average effort change of 0.39 for money-motivated participants. This would translate into a perceived-net-wage elasticity of effort change of 0.72. The comparison is less pronounced relative to the average overall effort of 9 sequences,

corresponding to an elasticity of 0.04. Are these large or small labor supply responses to withholding? We read these elasticities as substantial. Participants in our experiment do not respond very strongly to real economic incentives and a similar picture is true for wage earners in the real world[27]. To evaluate the economic significance of withholding we may be use a comparison with effects of changing the effective net wage. Reducing the gross payment per task from 50 to 30 cents (equivalent to a 40% effective tax) leads to an effort reduction of money-motivated participants (relative to the trend) of 0.21 sequences (MM6), translating into an elasticity of 0.06 relative to total effort. Hence, the perceived-net-wage elasticity is two thirds of the effective-net-wage elasticity, which we regard as an economically large effect.

While effects are substantial for our money-motivated participants, effects in the full sample are smaller and essentially zero for non-money-motivated participants. Our results therefore need to be qualified: To the degree that economic incentives matter, withholding tax rates significantly affect labor supply. Hence, participants who respond to monetary incentives also respond to withholding tax rates, although these do not matter for net wages.

## 5.2 Policy implications

Our results imply that, from a policy perspective, there is a trade-off between high perceived tax rates in some months and low perceived tax rates in others. Lower withholding rates reduce the perceived tax rates throughout the

[27] E.g., Bargain et al. (2014) estimate own-wage labor supply elasticities of European married men to be between 0.05 and 0.15.

year, but increase them at the end of the year due to larger tax adjustments. This trade-off sheds light on recent work suggesting that shrouded tax incentives may be welfare-enhancing since they mitigate potentially negative responses to taxation (Abeler and Jäger, 2015; Chetty et al., 2009). Our study demonstrates that mistaken responses to shrouded taxes will at some point become visible, be it in the form of large additional tax payments or refunds, and lead to a revision in behavior.

This raises the question how withholding taxes could be optimized and what could be their optimal economic salience (Goldin, 2015) in terms of household behavior (Sitzia et al., 2012). To answer this question, additional effects of withholding have to be considered, for example on saving, consumption or tax compliance. As pointed out in the introduction, many studies come to the conclusion that high withholding rates are desirable. Our results may be interpreted as an argument against high rates since it highlights the potential cost of lower perceived net wages throughout the year.

Our results may also be informative for the taxation of married couples. With joint taxation and progressive income tax rates, choosing adequate withholding rates is more difficult, since the effective tax depends on the income of both spouses. The withholding rate on the second earner's income cannot be computed without information on the first earner's income. Since employers usually do not have this information, withholding rates are necessarily incorrect in many cases. This is true for the United States and Germany where tax authorities differentiate withholding taxes for single and married individuals, or for the United Kingdom, Ireland and Australia where governments have implemented comprehensive withholding systems that col-

lect all taxes on wage income.[28] An undesirable effect of joint taxation may be that it increases the perceived marginal tax rates for the second earner and negatively affects labor supply at the intensive (by adjusting the number of hours worked) and the extensive margins (by quitting the labor market or opting for jobs in which performance above a certain minimum is not further rewarded). Our results indicate that both effects could arise and may lead to a higher fraction of second earners working less or opting out of performance-based jobs due to high withholding.

### 5.3 External validity

An obvious caveat of laboratory results is that it is difficult to translate them to the real world. However, the laboratory allows focusing on specific features of real-world labor markets and their effect on individual behavior. In this regard, we consider our experiment a proof of concept, testing for withholding rate effects when real-world complexities (time discounting, risk, tax law complexity, etc.) can either be abstracted from or be controlled for. Our results show that behavior is affected by theoretically irrelevant withholding taxes although participants are thoroughly explained how the tax system works. In real-world labor markets and in a more complicated tax system, without explanations provided and a large time gap between working and receiving tax refunds, effects similar to ours are likely. In this

[28] See for the US the tables in www.irs.gov/pub/irs-pdf/p15.pdf (p. 43ff), for Germany §§ 38, 38a-b, 39, 39a-f EStG (http://www.gesetze-im-internet.de/estg/BJNR010050934.html), for the UK a pay-as-you-earn system (http://www.hmrc.gov.uk/payerti/paying/deadline.htm), for Australia a pay-as-you-go system http://www.ato.gov.au/Business/Activity-statements/Pay-as-you-go-%28PAYG%29-income-tax-instalment/.

sense, our contribution provides first hints *how* withholding taxes may affect labor supply, but not *how much*. For the same reason, the fact that, in aggregate, our withholding and tax adjustment effects balance each other out does not extrapolate to the real world.

The results in Tables 5 and 6 demonstrate that the behavioral effects of withholding taxes, at least through tax adjustments, can be shown for the full sample, but are driven by money-motivated individuals. We do not consider this as a major limitation, since heterogeneity in work motivation may also be expected to be a feature of real-world labor markets. Workers who are mainly motivated by ambition, follow perceived work norms or simply enjoy work may not react strongly to a change in their extrinsic incentive environment. We therefore summarize that, as long as there are wage earners who are motivated by money, withholding taxes can have distortive effects.

# 6 Conclusion

In our experiment, we investigate the effect of withholding taxes on labor supply decisions. We vary the level of withholding rates which are, by construction, irrelevant for the decision-makers' economic incentives. We find that withholding taxes decrease labor supply, at least when decision makers are motivated by monetary incentives. Provided that our results can be interpreted in terms of real-world labor markets, policy-makers should pay attention to the choice of withholding taxes and the payment of refunds.

## References


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

Avram, S. (2015). Benefit losses loom larger than taxes: The effects of framing and loss aversion on behavioural responses to taxes and benefits. Technical report, Institute for Social and Economic Research.

Bargain, O., Orsini, K., and Peichl, A. (2014). Comparing labor supply elasticities in Europe and the United States: New results. *Journal of Human Resources*, 49(3):723–838.

Blumkin, T., Ruffle, B. J., and Ganun, Y. (2012). Are income and consumption taxes ever really equivalent? evidence from a real-effort experiment with real goods. *European Economic Review*, 56(6):1200–1219.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2010). Salience theory of choice under risk. Technical report, National Bureau of Economic Research.

Cabral, M. and Hoxby, C. (2012). The hated property tax: salience, tax rates, and tax revolts. Technical report, National Bureau of Economic Research.

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

Cheremukhin, A., Popova, A., and Tutino, A. (2011). Experimental evidence

on rational inattention. *Federal Reserve Bank of Dallas Working Paper*, 1112.

Chetty, R., Friedman, J. N., and 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.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–1177.

Chetty, R. and Saez, E. (2013). Teaching the tax code: Earnings responses to an experiment with eitc recipients. *American Economic Journal: Applied Economics*, 5(1):1–31.

Chow, C. W. (1983). 1983 competitive manuscript award: The effects of job standard tightness and compensation scheme on performance: An exploration of linkages. *Accounting Review*, pages 667–685.

Eliaz, K. and Spiegler, R. (2011). Consideration sets and competitive marketing. *Review of Economic Studies*, 78(1):235–262.

Engström, P., Nordblom, K., Ohlsson, H., and Persson, A. (2015). Tax compliance and loss aversion. *American Economic Journal: Economic Policy*, 7(4):132–164.

Erkal, N., Gangadharan, L., and Nikiforakis, N. (2011). Relative earnings and giving in a real-effort experiment. *American Economic Review*, pages 3330–3348.

Feldman, N. E. (2010). Mental accounting effects of income tax shifting. *Review of Economics and Statistics*, 92(1):70–86.

Feldman, N. E. and Ruffle, B. J. (2013). The impact of tax exclusive and inclusive prices on demand. *Monaster Center for Economic Research Discussion Paper No. 13-02*.

Finkelstein, A. (2009). E-ztax: Tax salience and tax rates. *Quarterly Journal of Economics*, 3(124):969–1010.

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

Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *The American Economic Review*, 100(1):541–556.

Fochmann, M. and Weimann, J. (2013). The effects of tax salience and tax experience on individual work efforts in a framed field experiment. *FinanzArchiv: Public Finance Analysis*, 69(4):511–542.

Fochmann, M., Weimann, J., Blaufus, K., Hundsdoerfer, J., and Kiesewetter, D. (2013). Net wage illusion in a real-effort experiment*. *Scandinavian Journal of Economics*, 115(2):476–484.

Gabaix, X. (2011). A sparsity-based model of bounded rationality. Technical report, National Bureau of Economic Research.

Gamage, D. and Shanske, D. (2011). Three essays on tax salience: Market salience and political salience. *Tax Law Review*, 65:19–98.

Goldin, J. (2015). Optimal tax salience. *Journal of Public Economics*, 131:115–123.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, pages 1–12.

Hayashi, A. T., Nakamura, B. K., and Gamage, D. (2012). Experimental evidence of tax salience and the labor-leisure decision: Anchoring, tax aversion, or complexity? *Public Finance Review*, pages 203–226.

Jones, D. (2012). Inertia and overwithholding: Explaining the prevalence of income tax refunds. *American Economic Journal: Economic Policy*, 4(1):158–185.

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

Kleven, H. J. and Kopczuk, W. (2011). Transfer program complexity and the take-up of social benefits. *American Economic Journal: Economic Policy*, pages 54–90.

Kőszegi, B. and Szeidl, A. (2013). A model of focusing in economic choice. *Quarterly Journal of Economics*, 128(1):53–104.

Kurzban, R. and Houser, D. (2005). Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations.

*Proceedings of the National Academy of Sciences of the United States of America*, 102(5):1803–1807.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of economic studies*, pages 175–208.

Rees-Jones, A. (2014). Loss aversion motivates tax sheltering: Evidence from us tax returns. *SSRN Working Paper 2330980*.

Reis, R. (2006). Inattentive consumers. *Journal of Monetary Economics*, 53(8):1761–1800.

Shapiro, M. D. and Slemrod, J. (1995). Consumer response to the timing of income: Evidence from a change in tax withholding. *American Economic Review*, pages 274–283.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Sitzia, S., Zheng, J., and Zizzo, D. J. (2012). Complexity and smart nudges with inattentive consumers. *SSRN Working Paper 2180956*.

Slemrod, J. B., Collins, B., Hoopes, J. L., Reck, D. H., and Sebastiani, M. (2014). Does credit-card information reporting improve small-business tax compliance? *SSRN Working Paper 2515630*.