Tax aversion in labor supply☆

Judd B. Kessler [a],[*], Michael I. Norton [b]

[a] *University of Pennsylvania, United States*
[b] *Harvard University, United States*

A B S T R A C T

In a real-effort laboratory experiment, labor supply decreases more with the introduction of a tax than with a financially equivalent drop in wages. This "tax aversion" is large in magnitude: when we decompose the productivity decrease that arises from taxation, we estimate that 40% is due to the lower net wage and the remaining 60% to tax aversion. This tax aversion affects labor supply more on the extensive margin (working less) than on the intensive margin (being less productive while working). The aversion is equally strong whether tax revenue goes to the U.S. government or back to the experimenter (a "laboratory tax"). We discuss the implications of our results for the relationship between labor supply and taxation.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

One of the most fundamental questions in public economics is the manner in which labor supply responds to taxation.[1] The effect of a tax on labor supply is assumed to work through the impact of the tax on net (i.e., after-tax) wages: when the tax rate increases, workers substitute away from labor and toward leisure because their net wage has fallen. A standard assumption is that the type of tax does not alter the labor supply response.

Recent research has demonstrated, however, that the relationship between taxation and labor is more complex; additional factors have been shown to mediate the effect of taxes on behavior. Field results on tax salience demonstrate that individuals respond less to a tax when the tax is less explicit (Chetty et al., 2009; Finkelstein, 2009; Cabral and Hoxby, 2012; Jones, 2012). Recent laboratory research has supported the claim that individuals respond more to taxes that are more salient (e.g., Sausgruber and Tyran, 2005; Blumkin et al., 2012), and increases in income taxes are often less salient than equivalent decreases in wages. Fochmann et al. (2013) found that laboratory workers who were paid the same net wage worked harder

[1] The elasticity of labor supply has been shown to be a sufficient statistic for identifying the income tax schedule that maximizes welfare (with regard to the equity-efficiency tradeoff in Mirrlees, 1971; see Saez, 2001 and Chetty, 2009).

and longer when they faced higher gross wages and higher tax rates: a result they called "net wage illusion" since workers behaved as though their net wage was higher because their gross wage was higher.

We highlight another channel through which taxes affect labor supply. We find that when taxes are just as salient as wages, individuals' aversion to taxes per se leads them to decrease their labor supply more in response to a new tax on labor income than to a financially equivalent decrease in wages. In a real-effort laboratory experiment, we saliently shock wages either by (a) directly decreasing gross wages or (b) introducing a tax on labor income. While the net (i.e., after-tax) wage is the same in these treatments, we find that subjects are significantly more likely to withhold effort — and produce significantly less — when the decrease in net wage is presented as a tax rather than a pay cut.[2]

We ground our predictions in the literature in psychology on aversion to taxation (e.g., McCaffery and Baron, 2006; Reimers, 2009). In Sussman and Olivola (2011), individuals report being more likely to travel further, and wait in line longer, when doing so avoids a tax rather than earning them a discount or avoiding some other cost.[3] Research has identified several mechanisms by which taxes exert a particularly negative effect. First, tax payments are decoupled from the public goods received by taxpayers, reducing taxpayers' perceptions that their taxes have provided them with any benefit (e.g., Mettler, 2011). Second, taxpayers frequently experience a lack of agency or influence over how their taxes are spent, further increasing tax aversion (Alm et al., 1999; Lamberton, 2013). We extend this literature by demonstrating that the aversion to taxes affects labor supply decisions in a real effort task. We also benchmark the magnitude of the effect against a direct drop in wage.

Our results also relate to a large experimental economics literature on taxation. Much of this literature examines the effect of taxes, which may or may not be salient and may or may not be complex, on workers who must calculate their net wage from their gross wage and the tax schedule (e.g., Swenson, 1988; Blumkin et al., 2012; Hayashi et al., 2013; Fochmann and Weimann, 2013; Fochmann et al., 2013; Weber and Schram, 2013; Doerrenberg and Duncan, 2014; Abeler and Jager, 2015).[4] Other research examines labor supply responses to net-of-tax wages without explicit mention of taxes (Sillamaa, 1998, 1999a,b). We add to this literature by identifying tax aversion as a motivator for labor supply. In addition, we do so in a setting where both gross and net wages are clearly and simply communicated to subjects, minimizing issues of complexity and salience that influence results in many of the previous experiments.[5]

Finally, our results speak to the elasticity of labor supply with respect to after-tax wages and the potential relevance of statutory incidence. Elasticities based on changes in tax policy may differ depending on whether the net wage is explicitly a result of taxation that is imposed on the worker (e.g., income taxes) or a result of taxation that is imposed on the firm (e.g., payroll taxes).[6] In particular, our results suggest we may observe elasticities larger in magnitude when income taxes are imposed directly on workers (so the worker is likely to interpret their decrease in wages as a tax) as opposed to payroll taxes imposed on employers that leads to an equivalent decrease in net wages (but where workers may be less likely to interpret the change as due to taxation, since they subjectively experience the change as a pay cut rather than a tax increase). This is consistent with evidence in Lehmann et al. (2013), which estimates in field data an elasticity near zero for changes in payroll taxes and a larger in magnitude (and significantly different from zero) elasticity for changes in income taxes.[7]

The paper proceeds as follows. Section 2 describes the experimental design. Section 3 presents the results. Section 4 discusses the results and concludes.

2. Experimental design

325 subjects participated in one of 33 sessions at the Computer Lab for Economic Research at Harvard Business School between April and August of 2012. Subjects were paid a $10 show-up fee for participating in the study and had the opportunity to earn extra money by working at a real effort task.

The task involved typing strings of nine characters in a specific order into a text entry box. Each nine-character string was made up of a combination of seven unique characters: / \ ) (⌐ < and the space bar. Subjects in the study performed the task for a total of 20 minutes in two 10-minute blocks. In each 10-minute block, every time a subject correctly typed a

---

[2] Unlike in the majority of the salience literature, net wages are explicitly communicated to subjects in all our treatments: regardless of whether the wage is cut directly or through a tax, subjects observe their cumulative net earnings as they work.

[3] There is evidence that tax aversion effects vary by political party. Hardisty et al. (2010) find that Republicans and Independents, but not Democrats, report being more willing to pay higher prices for environmentally friendly products when the higher cost is due to a "carbon offset" rather than a "carbon tax."

[4] Empirical studies on field data have similarly demonstrated that tax complexity leads taxable income to respond not to marginal tax rate but to average tax rate (Liebman and Zeckhauser, 2004; Feldman and Katuscak, 2006).

[5] Hayashi et al. (2013) failed to find tax aversion in the binary decision of whether to accept a job in the laboratory; when subjects were shown net wages alongside gross wages, labor supply did not vary with tax structure.

[6] To our knowledge, among the experimental literature within economics only Weber and Schram (2013) investigate statutory incidence — measuring subjects' effort when facing taxes imposed either on employees or on employers. In the experimental instructions used by Weber and Schram (2013), however, the word tax is never mentioned and the wage changes as workers provide effort, so the differences in productivity across treatments (observed among men but not women) appear to be due to their ability (or inability) to do the somewhat complicated math required to calculate the net wage.

[7] The explanation presented in Lehmann et al. (2013) centers on wage rigidity: that wages change less in response to employer side taxes as in Muysken et al. (1999). Our results support a not-mutually-exclusive alternative explanation based on tax aversion.
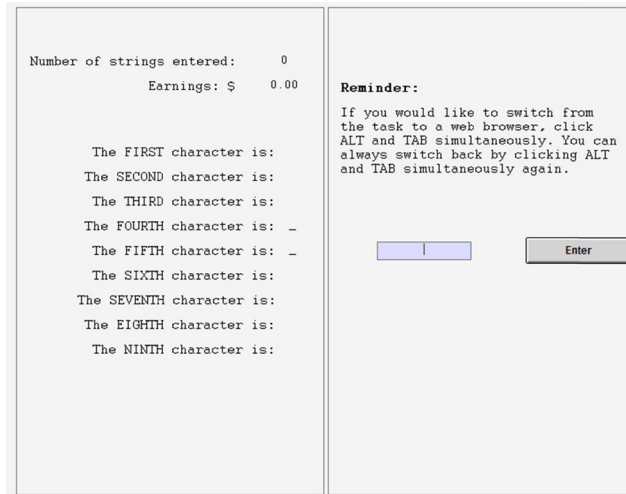
Fig. 1. Text entry screen. This text entry screen displays the interface subjects used in the experiment. The left panel showed the strings to be typed and the order of characters. The text entry interface prevented subjects from copying and pasting the string of characters and guaranteed that typing the characters as instructed was the fastest way to complete the task. The particular string shown in figure is three spaces, followed by two underscores, followed by four spaces or: "_". Once subjects typed the string and clicked the button labeled "Enter", the left panel would update the cumulative "Number of strings entered" to 1 and "Earnings" to $0.02 and a new set of nine characters would appear. Note that on the right panel of the text entry interface was a reminder to subjects that they could switch to using a web browser that was also open on the lab computer.

string of characters and pressed the enter button, a new string would appear until the subject ran out of time (see Fig. 1 for a picture of the text entry screen for the first 10-minute block).

To mirror real-world decisions, during the text entry task subjects could choose to forgo earning by switching to leisure. In particular, they were given the explicit option to surf the Internet.[8] At the start of the study, all subjects completed four test strings so that they understood how to perform the text entry task.

For the first ten-minute block, all subjects were paid a wage of $0.02 per string that they completed correctly. Subjects had the opportunity to complete as many strings as possible in the 10 minutes allotted. Subjects in all treatments completed this task for the same $0.02 wage (without any mention of a tax), so that performance in the first block could be used as a control for subjects' ability and interest in the task.

2.1. Wage and tax manipulations

Before the second ten-minute block, we experimentally varied whether subjects received a wage cut and, if so, whether the wage cut was associated with a tax. Subjects were randomly assigned to one of four wage and tax treatments that would apply to the second ten-minute block of work.

In the *$0.02 Treatment* (71 subjects), subjects were told they would continue to receive the same $0.02 wage per string as in the first 10-minute block. The instructions read (emphasis is in the original throughout):

> For each string that you complete in the second **10 minutes** of the task, you will earn **$0.02**.

In the *Wage Cut Treatment* (81 subjects), subjects were told they would receive a $0.01 wage per string. The instructions read:

> For each string that you complete in the second **10 minutes** of the task, you will earn **$0.01**.

Other than $0.01 replacing $0.02, the instructions are identical to the *$0.02 Treatment.*

In the *Experimenter Tax Treatment* (88 subjects), subjects were told that they would receive the same $0.02 wage per string but they would pay a 50% tax back to the experimenter for use in future studies. The instructions read:

> For each string that you complete in the second **10 minutes** of the task, you will earn **$0.02**. However, for each string that you complete in the second **10 minutes** of the task, you will face a **tax of 50%**. For each string you complete in the second **10 minutes** of the task, **$0.01** of your earnings will be taken from you and given to the experimenter for use in future studies.

---

[8] As an alternative to completing strings, subjects were provided with an unrestricted web browser to check email or visit other websites. Since we are interested in studying labor supply, we provided subjects with an alternative to working on the task that we thought would be desirable. Subjects may have also engaged in leisure in other ways (e.g., by reading material they brought with them into the lab or by silently using their cell phones).

Since the experimenter is paying the wage in all the treatments, the *Experimenter Tax Treatment* is just a framing manipulation of the *Wage Cut Treatment* with the lower wage being implemented by a financially equivalent tax.

In the *Government Tax Treatment* (85 subjects), subjects were told that they would receive the same $0.02 wage per string but they would pay a 50% tax to the federal government (i.e., half their earnings would be given to the U.S. Treasury). The instructions read:

> For each string that you complete in the second **10 minutes** of the task, you will earn **$0.02**. However, for each string that you complete in the second **10 minutes** of the task, you will face a **tax of 50%**. For each string you complete in the second **10 minutes** of the task, **$0.01** of your earnings will be taken from you and given to the United States Treasury for use by the U.S. Government.

All three *Wage Cut*, *Experimenter Tax*, and *Government Tax Treatments* provide workers a net wage of $0.01 per string. (None of the treatments contained an explanation for the decrease in net wages from $0.02 to $0.01.) As noted above, since the experimenter pays the wages in all treatments, the *Wage Cut* and *Experimenter Tax* treatments differ only in their framing. The *Government Tax* leads to the same net wage but each string also generates a $0.01 transfer from the experimenter to the federal government.[9]

The $*0.02 Treatment* has a different net wage than the other three treatments, but allows us to measure — by comparing productivity in the $*0.02 Treatment* to the *Wage Cut Treatment* — the effect on productivity of a 50% wage cut without a tax frame. With this magnitude, we can measure the extent of any tax aversion response to the tax frame in terms of the effect of cutting the wage alone.

After we exposed subjects to their wage and tax treatment, we allowed subjects to set a cap on their own productivity for the second block of 10 minutes. Setting this cap self-imposed a maximum number of strings that subjects could complete, serving the role of a self-commitment device used as a reaction to different wage structures (akin to an employee making a decision to cut back on the number of shifts or days she works, switch from full-time to part-time, or retire early). All subjects were asked to set a cap, which could be any number between 0 and 200, with 200 being the highest number of strings possible in the second block. Subjects who did not wish to constrain their productivity could simply set the cap to 200. Subjects then worked on the task for another 10 minutes or until they reached their cap, whichever came first. At the end of the study, subjects answered questions about the task and completed a demographic questionnaire.

In addition to observing a tax aversion response if it is present, we are interested in whether the tax aversion response arises through decisions about how much time to spend working or through decisions about how much effort to provide while working. We can separate these two channels using the timestamp data associated with completed strings. Note that how much effort to exert while working is clearly an intensive margin labor supply decision. How much time to spend working, however, can be interpreted as an intensive margin labor supply decision (e.g., if taking long breaks or setting a low cap reflect another form of work slowdown) or an extensive margin labor supply decision (e.g., if those same long breaks or low cap suggest an explicit decision not to work for some of the 10-minute window). Because long breaks and low caps are the only way for subjects to make extensive margin decisions in our laboratory setting, we take the latter interpretation. We use the timestamp data to create a measure of time spent working and treat this decision as the extensive margin, and we use speed while working — namely, the average number of seconds it takes to complete a string — as the intensive margin.

2.2. Meaning manipulation

While the primary focus of the study is on the effect of taxes on labor supply, we hypothesized that any effect of taxation might be mitigated if the subjects felt that the task that they were doing had some meaning (Ariely et al., 2008). We therefore crossed our four wage and tax treatments with variation in the meaning of the task, creating a 4 × 2 factorial design. Table A1 in the Appendix reports the number of subjects in each of these 8 cells.

We selected the specific strings subjects typed in sets of four. Subjects in the *Control Treatment* (170 subjects) completed each set of four strings in a random order; subjects in the *Meaning Treatment* (155 subjects) completed the set of four strings in a slightly different order. The order of strings in the *Meaning Treatment* was chosen so that when the strings were presented vertically on the screen they would display a series of bubble letters, which in the aggregate spelled out the titles of Beatles songs. Subjects saw their completed strings from the first block for 90 seconds before they began the second block of 10 minutes. Fig. A1 in the Appendix shows the screens.

While subjects in the *Meaning Treatment* were more likely to find meaning in their work, the difference in productivity between the *Meaning* and *Control* treatments is never significant in any specification, neither across all treatments nor within any of our wage and tax treatments; we summarize these results in the Appendix. Consequently, we combine the data from the *Meaning* and *Control* treatments in the analyses that follow.

---

[9] In both the *Experimenter Tax* and *Government Tax* treatments, the collected tax revenue was used as indicated to subjects: revenue was returned to the experimenter's research budget for use in future studies or sent to the U.S. Treasury (see http://www.fms.treas.gov/faq/moretopics_gifts.html), respectively.
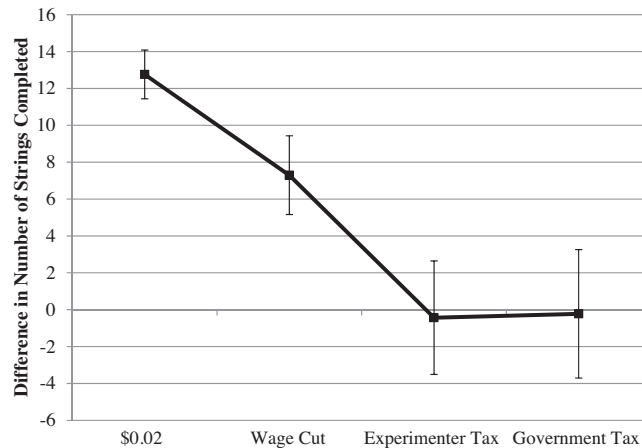
Fig. 2. Difference in number of strings completed. Figure shows the average difference in number of strings completed between the first and second block for each wage and tax treatment. Standard error bars are shown around each mean. For subjects in the *$0.02 Treatment*, for whom wage does not change between block one and block two, the average difference in productivity is an increase of 12.76 strings; for subjects in the *Wage Cut Treatment*, the average difference is 7.30 strings; for subjects in the *Experimenter Tax Treatment*, the average difference is −0.43 strings; for subjects in the *Government Tax Treatment*, the average difference is −0.22 strings.

3. Results

Subjects completed the first 10-minute block before being randomized to treatment and, as expected, there was no difference between treatments in the average number of strings completed in the first 10-minute block (see Table A2 in the Appendix for summary statistics by treatment; see the "Pre-Treatment Productivity" panel of Fig. A2 in the Appendix for CDFs of strings completed by treatment).[10] However, there was a great deal of individual heterogeneity in the number of strings individuals completed in the first 10-minute block, n_1, which suggests that productivity in the first block may be a useful control for ability and interest in the typing task. Consequently, in all specifications we either control for a subject's n_1 or investigate the difference, d, in the number of strings a subject completed in block two, n_2, and in block one, n_1 (i.e., the difference in productivity between the blocks is $d = n_2 - n_1$).

The most straightforward way to visually show the results from manipulating wage and tax is to report the difference in productivity, d, for each of the four wage and tax treatments. Fig. 2 shows these differences, and three results are apparent.

First, the 50% wage cut decreases productivity in the second block: the improvement from the first to second block is smaller (7.30 strings) in the *Wage Cut Treatment* than in the *$0.02 Treatment* (12.76 strings).[11]

Second, there is an additional, large decrease in productivity associated with simply framing the same $0.01 decrease in wage as a tax rather than a wage cut. The average change in productivity in the *Experimenter Tax Treatment* is directionally negative (−0.43 strings).

Third, the effect of framing the wage cut as a tax is similar in magnitude regardless of whether the tax is paid to the government or back to the experimenter. The average change in productivity in the *Government Tax Treatment* (−0.22 strings) is very similar to that in the *Experimenter Tax Treatment*. These results demonstrate "tax aversion": a decrease in productivity due simply to the fact that the same net decrease in wages is being implemented as a tax.

Table 1 presents regression results highlighting these three findings. Regressions (1) through (3) have as a dependent variable the difference in productivity going from the first to second block of 10 minutes while regressions (4) through (6) have as a dependent variable the number of strings completed in the second block, controlling linearly for the number of strings completed in the first block of 10 minutes.[12]

The first regression in each set — (1) and (4) — investigates the impact of a 50% wage cut on productivity in the second block, comparing the productivity of subjects in the *$0.02 Treatment* to subjects in all three other treatments with a net wage of $0.01. There is a significant decrease in productivity among subjects who are paid $0.01 rather than $0.02 in the second block.

---

[10] As expected by random assignment, there were also no differences in the demographic make-up of the groups randomly assigned to each treatment (see Table A3 in the Appendix). As one would expect from random assignment and no differences in demographics across treatments, including demographics as controls do not affect the main results of the paper (see Table A4 in the Appendix, which replicates Table 1, below, but includes demographic controls).

[11] Note that d is generally positive: most subjects are capable of improving from the first to second 10-minute block. In particular, in the *$0.02 Treatment* — when the wage stays the same across blocks — 93% of subjects completed more strings in the second block than in the first; of the 7% who did not, 4 out of 5 set a cap below their performance in the first block.

[12] Regressions (1) through (3) are very similar to regression (4) through (6) but have the coefficient on *Strings in first block* set to 1 rather than fit to the data.

Table 1
Effect of the net wage change and the tax frame on productivity.

	Difference in productivity (strings completed in second block – strings completed in first block)			Strings completed in second block		
	(1)	(2)	(3)	(4)	(5)	(6)
Net wage of $0.01	−10.66 (2.18)***	−5.46 (2.51)**	−5.46 (2.51)**	−10.67 (2.15)***	−5.40 (2.48)**	−5.40 (2.48)**
Tax		−7.62 (3.15)**	−7.73 (3.75)**		−7.74 (3.12)**	−7.53 (3.72)**
U.S. Treasury			0.21 (4.65)			−0.43 (4.60)
Strings in first block				1.13 (0.05)***	1.13 (0.05)***	1.13 (0.05)***
Constant	12.76 (1.32)***	12.76 (1.32)***	12.76 (1.32)***	0.671 (4.61)	0.463 (4.49)	0.412 (4.41)
Observations	325	325	325	325	325	325
R-squared	0.03	0.05	0.05	0.52	0.52	0.52

Table shows OLS regressions reporting the difference in productivity between the first and second block — regressions (1) through (3) — and the productivity in the second block — regressions (4) through (6). The constant shows the value for the *$0.02 Treatment*. *Net wage of $0.01* indicates that the subject was paid a net wage of $0.01 rather than $0.02. *Tax* indicates that the net wage of $0.01 was implemented using a 50% tax. *U.S. Treasury* indicates that revenue from the tax imposed was sent to the U.S. Treasury rather than paid back to the experimenter. *Strings in first block* is the number of strings the subject completed in the first ten-minute block. Robust standard errors are in parentheses.
* Significance at 0.1.
** Significance at 0.05.
*** Significance at 0.01.

The second regression in each set − (2) and (5) − shows that productivity decreases more when the 50% net wage cut is implemented through a tax. Here, the coefficient on *Net wage of $0.01* reflects the difference in productivity between the *$0.02 Treatment* and the *Wage Cut Treatment*, which is a statistically significant difference of about 5.5 strings.[13] The coefficient on *Tax* reflects the difference in productivity between the *Wage Cut Treatment* and the two tax treatments. This difference in productivity is statistically significant and between 7.5 and 8 strings, demonstrating tax aversion: given the same net wage, subjects produce less when they are taxed.

The third regression in each set − (3) and (6) − splits the two tax treatments. While the coefficient on *Net wage of $0.01* does not change (it is still the difference between the *$0.02 Treatment* and the *Wage Cut Treatment*), here the *Tax* coefficient reflects the difference in productivity between the *Wage Cut Treatment* and the *Experimenter Tax Treatment*, demonstrating that tax aversion arises even in a pure framing manipulation.

Finally the coefficient on *U.S. Treasury* reports the difference in productivity between the *Experimenter Tax Treatment* and *Government Tax Treatment*. That this coefficient is small in magnitude and not statistically significant from 0 demonstrates that subjects are equally tax averse whether the tax revenue is being returned to their employer (in this case the experimenter) or sent to the U.S. government.

The latter two specifications in each set of regressions allow us to benchmark the magnitude of the tax aversion effect against the effect of directly cutting the wage from $0.02 to $0.01. The tax aversion effect (i.e., the productivity drop due solely to the fact that the same net wage cut was induced by a tax) is around 7.6 strings, about 140% of the wage cut effect of around 5.4 strings. This signifies that when a 50% tax is implemented in our experiment, about 40% of the resulting decrease in productivity is due to the lower net wage while 60% of the decrease is due simply to its being labeled as a tax.

3.1. Choosing not to work or producing less while working?

One natural question to ask is whether the tax aversion effect on labor supply arises due to individuals either: (a) working less (akin to working fewer days or fewer hours each day), which we treat as an extensive margin decision, and/or (b) providing less effort while on the job, which we treat as an intensive margin decision.

Subjects in our experiment were explicitly given the opportunity to engage in leisure by switching away from the zTree work interface to a web browser that was open and connected to the Internet on their computer terminals.[14] Our approach to identify subjects' extensive and intensive margin labor supply decisions is to treat the extensive margin decision as how much time subjects spent engaged with the zTree interface working toward completing strings — rather than surfing the Internet or doing some other activity — and the intensive margin decision as how fast they completed strings while engaged in the task.

Subject productivity data is time-stamped (i.e., we know when each string is entered), which allows us to determine when subjects are actively completing strings and when they are taking a break or have stopped working completely.[15]

---

[13] That a wage cut has a negative effect on labor supply has also been shown in field settings (see e.g., Kube et al., 2013).

[14] In addition, subjects had the implicit opportunity to: read material they brought with them into the lab, silently use their cellphones, do their homework, or engage in any other silent activity during the lab session (i.e., they were not explicitly encouraged or discouraged from any of these activities).

[15] We do not observe exactly when subjects are using the web browser, as they used keystrokes "alt" and "tab" to switch from the zTree interface to the web browser and back. Similarly, we do not observe exactly when subjects are using their cell phones, reading material they brought with them into the lab, or doing something else that draws their attention away from the work interface. However, we can observe exactly when subjects are completing strings in the zTree interface, and we infer that other time was spent engaged in "leisure" activities.

Table 2
Effect of the net wage change and the tax frame on time spent working.

	Difference in time spent working			Time spent working in second block		
	All time until last string (1)	Excluding breaks > max time (2)	Excluding breaks > 90th pctl (3)	All time until last string (4)	Excluding breaks > max time (5)	Excluding breaks > 90th pctl (6)
Net wage of $0.01	−25.28 (12.85)**	−26.87 (12.85)**	−22.00 (12.82)*	−25.26 (12.87)*	−26.87 (12.87)**	−22.11 (12.89)*
Tax	−40.35 (17.24)**	−44.37 (17.37)**	−47.41 (16.69)***	−40.36 (17.27)**	−44.37 (17.40)**	−47.30 (16.78)***
Time working in First Block				1.02 (0.06)***	1.00 (0.06)***	0.96 (0.07)***
Constant	−10.25 (4.19)**	−13.68 (4.28)***	−1.03 (5.20)	−20.35 (37.37)	−16.53 (37.08)	7.06 (37.70)
Observations	325	325	325	325	325	325
R-squared	0.04	0.05	0.05	0.21	0.21	0.22

Table shows OLS regressions reporting the difference in time spent working between the first and second block — regressions (1) through (3) — and the time spent working in the second block — regressions (4) through (6). The dependent variables are defined as follows: *All time until last string* measures time working until the subject completes his last string of the block. *Excluding breaks > max* subtracts from time spent working any gap in work longer than the longest amount of time it took a subject to complete a string in the first block of 10 minutes. *Excluding breaks > 90th pctl* subtracts from time spent working any gap in work longer than the 90th percentile of time it took a subject to complete a string in the first block of 10 minutes. The constant shows the value for the *$0.02 Treatment*. *Net wage of $0.01* indicates that the subject was paid a net wage of $0.01 rather than $0.02. *Tax* indicates that the net wage of $0.01 was implemented using a 50% tax (either to the experimenter or to the U.S. Treasury). Note that all of these regression specifications mirror the structure of regression specifications (2) and (5) in Table 1 where the coefficient on *Net Wage of $0.01* estimates the difference between the *$0.02 Treatment* and the *Wage Cut Treatment* and the *Tax* coefficient estimates the effect of implementing the wage cut as a tax. *Time working in First Block* is the time the subject spent working in the first ten-minute block by the same definition. Robust standard errors are in parentheses.

* Significance at 0.1.
** Significance at 0.05.
*** Significance at 0.01.

To measure extensive margin labor supply decisions, we identify when a subject stopped working by completing his or her final string and treat time before the final string as "time spent working" and any remaining time left in the 10-minute block as leisure. In alternative measures of extensive margin labor supply, we subtract from time spent working any long breaks between completed strings in which we can reasonably assume the subject spent some time explicitly choosing not to work (i.e., we look for gaps between completed strings long enough for the subject to have spent some time engaged in another activity before returning to the interface to complete the next string). We perform this latter analysis a number of times, reporting on two different thresholds for a long-enough gap between strings to constitute an explicit decision not to work.[16] Long breaks before subjects complete their last string are extremely rare (only 0.23% of strings in the second block are completed more than 30 seconds apart; only 0.004% are completed more than 60 seconds apart), and so changes in the definition of how long a break must be to include it as an explicit decision not to work has almost no effect on the estimates.

Table 2 shows this analysis. Since both tax treatments had an identical tax aversion effect in Table 1, we combine them in the analysis that follows.[17] In all regressions, by all definitions of time spent working, we see a significant and negative effect of being in one of the two tax treatments. In particular, *Tax* is negative and significant, demonstrating that subjects chose to work less on the extensive margin when the wage cut they experienced came in the form of a tax. Additionally, in all specifications we see a negative effect in labor supply on the extensive margin from the wage cut itself — *Wage Cut* is negative and at least marginally statistically significant for all tests.[18]

Regressions (1) and (4) treat all time until the subject correctly enters his or her last string as time spent working. By this measure, subjects spend on average 8 minutes and 58 seconds of the possible 10 minutes working (i.e., 89.6%). The other regressions estimate the same specification but subtract from the measure of working time any long breaks a subject takes before entering his or her last string. To identify when the subject is taking a long break, we use a subject's performance in the first block of 10 minutes to calculate a reasonable amount of time in which the subject could be expected to complete a string. Whenever a subject takes longer than this amount of time to complete a string, we say the subject is taking a long break and subtract from time spent working those excess seconds.[19]

[16] We use many other criteria for robustness checks, discussed in footnotes, and always get the same pattern of results.

[17] As expected, results do not change if we look at the *Experimenter Tax Treatment* and *Government Tax Treatment* separately. For each of the regressions shown in Table 2, we get a statistically significant tax aversion affect (at least $p < 0.1$, usually $p < 0.05$) for each of the two treatments separately (i.e., we get a negative effect comparing each of the treatments separately to the *Wage Cut Treatment*).

[18] As in Table 1, we control for individual heterogeneity in time spent working by subtracting the same measure for the subject in the first block of 10 minutes, as in Regressions (1) through (3), or controlling for it linearly, as in Regressions (4) through (6). However, this is not necessary to get the tax aversion effect. Regressing time spent working (under any of the definitions) on *Net Wage of $0.01* and *Tax*, the coefficient on *Tax* is significant with $p < 0.05$ even when *Time working in First Block* is not included as a control.

[19] For example, if the longest it took a subject to complete a string in the first block of 10 minutes was 21 seconds, then for a gap between strings of 1 minute, we would subtract $60 − 21 = 39$ seconds from working time. This method has the potential to account for breaks that last a few minutes (e.g., switching to the web browser to check email or read an article online) or a few seconds (e.g., checking a cell phone to respond to a text message).

Table 3
Effect of the net wage change and the tax frame on seconds per string.

	Difference in speed (seconds per string)			Speed in second block (seconds per string)		
	All time until last string (1)	Excluding breaks > max time (2)	Excluding breaks > 90th pctl (3)	All time until last string (4)	Excluding breaks > max time (5)	Excluding breaks > 90th pctl (6)
Net wage of $0.01	−0.347 (0.475)	−0.365 (0.473)	0.044 (0.121)	−0.015 (0.156)	0.005 (0.147)	0.099 (0.079)
Tax	0.841 (0.533)	0.567 (0.479)	0.059 (0.120)	0.559 (0.344)	0.253 (0.197)	0.022 (0.083)
Speed in First Block				0.527 (0.217)**	0.473 (0.167)***	0.787 (0.057)***
Constant	−0.909 (0.066)***	−0.946 (0.066)***	−0.775 (0.053)***	2.28 (1.46)	2.60 (1.13)**	0.565 (0.355)
Observations	317	317	317	317	317	317
R-squared	0.01	0.01	0.00	0.29	0.52	0.89

Table shows OLS regressions reporting the difference in speed (number of seconds per string completed) between the first and second block — regressions (1) through (3) — and the speed in the second block — regressions (4) through (6). The constant shows the value for the *$0.02 Treatment*. Note that the negative and significant effect on the constant in Regressions (1) through (3) demonstrate that subjects in the $0.02 treatment get faster at completing strings between the first and second blocks. *Net wage of $0.01* indicates that the subject was paid a net wage of $0.01 rather than $0.02. *Tax* indicates that the net wage of $0.01 was implemented using a 50% tax (either to the experimenter or to the U.S. Treasury). *All time until last string* measures speed using time working until the subject completes his last string of the block. *Excluding breaks > max time* subtracts from time working any gap in work longer than the longest amount of time it took a subject to complete a string in the first block of 10 minutes. *Excluding breaks > 90th pctl* subtracts from time working any gap in work longer than the 90th percentile of time it took a subject to complete a string in the first block of 10 minutes. *Speed in First Block* is the number of seconds per string the subject spent working in the first ten-minute block by the same definition. Robust standard errors are in parentheses.
* Significance at 0.1.
** Significance at 0.05.
*** Significance at 0.01.

In particular, the measure *Excluding breaks > max time* in Regressions (2) and (5) uses the longest amount of time it took a subject to complete a string in the first block of 10 minutes as a conservative estimate of the maximum length of time it should take the subject to complete a string. By this measure, 32% of subjects take at least one qualifying break and subjects spend on average 8 minutes and 51 seconds of the possible 10 minutes working (88.4%). That this measure is so close to the 8 minutes and 58 denoted above is a reflection of the fact that very few long breaks are taken. Indeed, only 8% of subjects take cumulative breaks of more than 20 seconds.

The measure *Excluding breaks > 90th pctl* in Regressions (3) and (6) uses the 90th percentile of the time that a subject took to complete strings in the first set of 10 minutes as the maximum amount of time it should take subjects to complete a string. The measure is more inclusive of potential breaks but possibly overestimates by counting as breaks times that subjects are just working slowly. By this measure, 95% of subjects are counted as having taken at least one break, and subjects are counted as having worked on average 8 minutes and 24 seconds (84.0%).[20] By all definitions of time spent working, we find a large and significant extensive margin labor supply response to the tax treatments.

The accompanying analysis for the intensive margin is shown below in Table 3. To perform this analysis, we construct a speed variable (i.e., the average number of seconds it takes to complete a string) for each subject for the first block of 10 minutes and the second block of 10 minutes. We construct this measure using each of the working time estimates used in Table 2. Since speed is the number of seconds it takes to complete a string, larger numbers mean subjects are working more slowly (note that this speed variable is not defined for subjects who do not complete any strings and they are dropped from the analysis). The question of whether workers respond negatively to the tax on the intensive margin asks whether this speed is slower in the tax treatments.

As can be seen in Table 3, there is no effect on the intensive margin in response to either the wage cut or that wage cut coming in the form of a tax in any specification. While the effect of tax is directionally positive, it is never statistically significant, even in Regressions (1) and (4), which are the most aggressive specifications in which any breaks before the last string is completed are treated as time spent working, potentially inflating estimates of the effect of labor supply on the intensive margin by miscounting (extensive margin) breaks as (intensive margin) work slowdowns.

The directional effect gets smaller in magnitude as our specifications get more conservative by subtracting off more breaks from the time spent working estimates. Finally, while the regressions control for individual heterogeneity in speed by subtracting seconds per string in the first block, as in Regressions (1) through (3), or controlling for it linearly, as in Regressions (4) through (6), this is not necessary for the results. When we drop *Speed in First Block* and rerun the specifications shown in Regressions (4) through (6) but without that variable, we never see a coefficient on *Tax* that is statistically significant (untabulated, *p* > 0.5 for all 3 tests).

---

[20] For robustness, we run the same specifications with other methods of determining when subjects are taking a break. We always get a strong tax aversion effect on labor supply on the extensive margin. For example, we get the same results setting the duration of how long it should take to complete a string at 21 seconds or 12 seconds (respectively the 99th and 95th percentile of time it took to complete strings in the first block, across all subjects) and estimate working time using that same duration for all subjects. We get the same pattern when using other measures for the threshold, including the 95th, 75th, or 50th percentile of strings in the first block, or setting the threshold based on a number of standard deviations greater than the subject's mean time to complete a string in the first block.

In addition to identifying that the tax aversion arises on the extensive labor supply margin of deciding how much to work, rather than the intensive margin of how productive to be while working, we can investigate whether subjects use the self-commitment device of caps on productivity disproportionately in response to the wage cut being framed as a tax. In the Appendix, we show analysis demonstrating that this is the case. Subjects are more likely to set a low cap that binds their productivity in the tax treatments than in the *Wage Cut Treatment*.

3.2. Who chooses not to work?

We have seen that this tax aversion in labor supply is a result of a subset of subjects choosing to withhold productivity by working for less time or by setting a low cap. Who are the people who display this aversion to taxation? We find no systematic difference of tax aversion (measured by its effect on productivity, the amount of time subjects spend working, or the likelihood of setting a low cap) based on performance in the first block or for any of the variables we collected in the demographic survey, including gender, age, student status, income level, liberal vs. conservative, or political affiliation. Interacting these demographics with being in a tax treatment in regressions on difference in strings completed, difference in time spent working, or probability of setting a low cap did not result in any significant interactions.[21] In summary, no characteristics of subjects that we observe before they were treated affect the likelihood that they display tax aversion in labor supply.

4. Discussion

We show that a net decrease in wages that is framed as a tax lead to less effort provision than the same wage decrease framed as a pay cut. Moreover, the tax aversion effect is large. When we decompose the decrease in productivity that comes from the introduction of a 50% tax, only 40% of the decrease is due to the lower net wage while 60% of the decrease is due to tax aversion. Put another way, the drop in productivity we observe would be less than half as large if the tax were implemented as a direct wage cut rather than as a tax. Furthermore, our analyses indicate that the effect of taxes on productivity is driven by people choosing to work less under the tax regime (which we interpret as an extensive margin labor supply decision) rather than being less productive while working (which we interpret as an intensive margin labor supply decision).

We additionally demonstrate that the effect is due to framing the wage cut as a tax rather than a distrust or dislike of the use of tax revenue by the U.S. government in particular. Subjects respond just as negatively when the tax revenue is returned to the experimenter (i.e., a direct framing manipulation from the wage cut, since wages are paid by the experimenter) as they do when the tax revenue collected is sent to the U.S. Treasury.

We note that our results may demonstrate an upper bound on the impact of tax on labor supply for three reasons. First, previous research demonstrates that taxes have a larger effect when more salient (e.g., Chetty et al., 2009; Blumkin et al., 2012), typical of a focusing illusion where making any attribute particularly salient leads that attitude to exert an outsized influence on decision-making (Kahneman et al., 2006). In our experiment, taxes are made explicit at the very moment that workers choose their labor supply decision. Consequently, the level of taxation is likely to be more salient than in many real-world settings (where taxes are likely not as salient on a day-to-day basis). Indeed, there is little evidence at the macroeconomic level that higher taxation reduced economic growth in the 20th century in the United States (Kuziemko et al., 2015). Second, our policy changes — the wage cut and the tax increase — are severe in magnitude: fully 50% of wages go back to the experimenter or to the government. Third, we introduce a tax in an environment where previously no taxes are imposed, and it is possible that the tax aversion effect is stronger when a tax is introduced than when the rate of a preexisting tax increases. Future work could help to tease apart when the tax aversion effect will be most pronounced and when it might be mitigated.

At the same time, it is quite clear that taxation appears to exert a particularly negative influence on labor supply in our experiment. Our results suggest a reason for the relevance of statutory incidence in labor taxation. Salient taxes imposed on workers may trigger tax aversion and lead people to withhold labor supply more than similar taxes levied on employers and experienced by employees as a change in wage. Tax aversion could thus lead to the kind of pattern observed in Lehmann et al. (2013), in which changes in payroll taxes are estimated to have a smaller effect on labor supply than equivalent changes in income taxes.

We also note that our paradigm allowed subjects to set a cap for their labor after learning about changes to their wages — in a sense giving them a self-commitment device to work less depending on their reaction to their wages. While we did not include a "no cap" treatment in our experiment, future research should explore the change in labor supply in response

---

[21] We conducted this analysis by: (a) creating at least one dummy variable based on each demographic variable (in particular, whether a subject was female, whether a subject was born before 1990, whether a subject had at least an associates or bachelors degree, whether a subject was currently a student, whether a subject had household income above $3000 a month, whether a subject self defined as liberal or conservative, and whether a subject was a Democrat), and based on first-block productivity (in particular, whether a subject completed more than 93 strings, the median in the first block) and (b) interacting this dummy variable with being in a tax treatment in a regression on either: difference in strings completed, difference in time spent working (for each of the three definitions in Tables 2 and 3), or probability of setting a low cap (for each of the definitions in Table A5). These regressions included all subjects in the *Wage Cut Treatment* and tax treatments for whom we had the relevant demographic data. The interaction effect was marginally statistically significantly different than 0 ($p < 0.1$) in only 1 of these 54 regressions.

to changes in taxation with and without the presence of a cap. We suggest that subjects' changes to their cap most closely simulate reactions of shift workers to changes in taxation by altering their number of shifts they take or days of work. For salaried employees, who do not typically have the option to select fewer shifts, setting a cap might mean switching from full-time to part-time or deciding to retire earlier than they might have otherwise.

Our subjects are drawn from a pool in the United States, raising the question of whether the tax aversion we observe here would hold in other countries with different taxation policies and attitudes toward taxation. For example, the World Values Survey includes an item asking respondents to report the extent to which cheating on taxes, if given the chance, is justified; using this data, Alm and Torgler (2006) report that the United States has higher tax morale (i.e., reports that tax evasion is less justified) than 16 European countries, including countries such as Norway, Sweden, and Denmark. In addition to attitudes toward tax compliance, attitudes toward increasing taxes are also likely to influence labor reactions to tax increases; within Europe, for example, citizens of Denmark, Sweden, and Finland are more supportive of tax increases than citizens of Ukraine and Russia (Svallfors, 2012). Examining reactions to tax increases of citizens from countries that vary in attitudes toward taxation offers a fruitful avenue for future research.

Can anything be done to mitigate tax aversion? Recent research suggests that two psychological barriers may be in part responsible for tax aversion: the decoupling of tax payments and the services citizens receive in return and a lack of agency in deciding how taxes are spent. To the first, interventions that "show the government's work" — for example, apps that reveal the progress that government in making in repairing potholes and replacing street signs — have been shown to increase support for government programs (Buell and Norton, 2011, 2013). To the second, giving taxpayers a voice in where their taxes are spent — introducing taxpayer agency by soliciting opinions on the ultimate targets of tax spending — reduces tax dodging and increases satisfaction with paying (Lamberton et al., 2014). Future research is needed to further understand workers' sensitivity to different forms of taxation, as well as how to mitigate the effect of tax aversion on labor supply.

Appendix A.

A.1. Meaning manipulation results

Our post-study survey demonstrated that subjects in the *Meaning* treatment recognized that their strings spelled words. Subjects in the *Meaning* treatment were much more likely to answer the multiple-choice question: "Did the strings you typed during the task mean anything?" with the answer: "Yes, they became words" (85.8% of subjects in *Meaning* vs. 1.8% of subjects in *Control*; OLS with robust standard errors, 325 obs, $p < 0.001$). Interestingly, over half the subjects in the control treatment chose to answer the multiple choice question: "Did the strings you typed during the task mean anything?" with the answer: "Yes, they became pictures" implying that they found meaning in their work even when it was intended to be meaningless. When combining these two affirmative answers to the meaning question, however, subjects were still more likely to find meaning in their work in the *Meaning* treatment (89.6% in *Meaning* vs. 54.1% of subjects in *Control*; OLS with robust standard errors, 325 obs, $p < 0.001$).

However, even though subjects in the *Meaning* treatment were more likely to report that the strings meant something, it did not affect productivity. The average difference in productivity from the first to second block was 3.41 strings in the *Control* treatment and 5.55 in the *Meaning* treatment (OLS with robust standard errors, 325 obs, $p = 0.439$). In addition, the difference is productivity between the *Meaning* and *Control* treatments is not significant within any wage and tax treatment
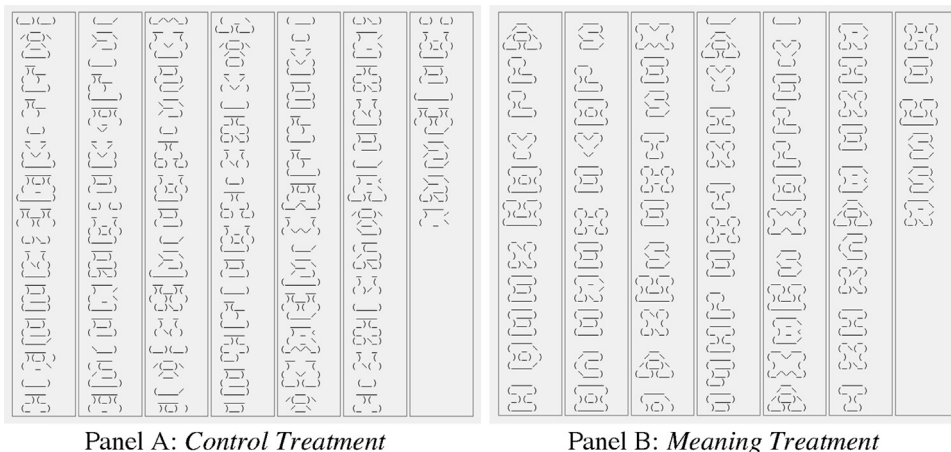

Panel A: *Control Treatment*   Panel B: *Meaning Treatment*

Fig. A1. Completed strings in first ten-minute block. Panel (A) shows what a subject who completed 307 strings in the *Control Treatment* would see for 90 seconds between the first and second 10-minute block of work. Fig. 2 panel (B) shows the equivalent picture for subjects in the *Meaning Treatment*. Subjects who completed fewer strings saw less text. Strings entered in the first 10-minute block were shown starting at the top of the leftmost column. Each column is 47 strings, so the median subject (who completed 93 strings in the first 10-minute block) saw two columns of text.

Table A1
Number of subjects in each treatment.

		Second work period pay treatment			
		$0.02	Wage Cut	Experimenter Tax	Government Tax
Meaning treatment	Control	35	40	56	39
	Meaning	36	41	32	46

Table shows the number of subjects in each of the 8 cells in the experiment. Subjects were randomly assigned to one of these 8 cells with equal probability using the random number generator in zTree.

Table A2
Summary statistics by treatment.

	Treatment			
	$0.02	Wage Cut	Experimenter Tax	Government Tax
obs	71	81	88	85
n_1	91.24 (20.68)	90.73 (24.26)	89.25 (23.36)	93.99 (19.95)
n_2	104 (25.35)	98.02 (33.03)	88.82 (39.88)	93.76 (38.95)
d	12.76 (11.15)	7.30 (19.20)	−0.43 (28.88)	−0.22 (32.12)
cap	167.03 (47.11)	164.25 (55.50)	144.82 (67.73)	153.68 (65.92)

Table shows summary statistics from the experiment with means and standard deviations in parentheses. *obs* is the number of subjects in each treatment. n_1 is the average number of strings completed in each treatment during the first 10-minute block. n_2 is the average number of strings completed in each treatment during the second 10-minute block. *d* is the average difference, $d = n_2 − n_1$ (note that all numbers are rounded to two decimal places). *cap* is the average cap set for the second 10-minute block. None of the differences across treatments in pre-treatment productivity, n_1, are statistically significantly different from 0 ($p > 0.1$ for all pairwise test, both *t*-tests and Kolmogorov–Smirnov tests).

Table A3
Demographics by treatment.

	Treatment			
	$0.02	Wage Cut	Experimenter Tax	Government Tax
obs	71	81	88	85
Female	45.1%	57.0%	54.7%	50.6%
Age	22 (3.99)	22.77 (4.61)	22.14 (3.50)	22.30 (3.47)
Student	80.0%	73.4%	80.2%	70.2%
Monthly Income	$2488 ($3260)	$3215 ($3700)	$2370 ($3156)	$2594 ($3254)
Liberal (vs. Conservative)	4.89 (1.53)	5.01 (1.45)	5.22 (1.36)	5.02 (1.54)
Democrat	46.5%	51.3%	51.8%	52.4%

Table shows demographics from the experiment with percentages or, alternatively, means and standard deviations in parentheses. *obs* is the number of subjects in each treatment. *Female* is the percentage of subjects who were female in each treatment. *Age* is the average age of subjects in each treatment (age is imputed by subtracting the year of birth from the year the study was run). *Student* is the percentage of subjects who were either a full-time or part-time student. *Monthly Income* is average monthly income as imputed from the 11 possible income choices by selecting the midpoint of the range of options (except for over $10,000, which is coded as $10,000). *Liberal (vs. Conservative)* is the average choice on the scale from 1 = "Very Conservative" to 7 = "Very Liberal" (with 4 = "Neither Conservative nor Liberal"). *Democrat* is the percentage of subjects who report that they identify as Democrat, rather than Republican, Independent, or Other. Non-responses for a demographic trait are dropped when computing these statistics. None of the differences between treatments are statistically significant ($p > 0.1$ for all pairwise tests, either tests of proportions or *t*-tests depending on whether the data is binary or continuous).

in any specification. This suggests that on average typing strings that become bubble letters to spell Beatles song titles is no more motivating for subjects than typing strings that became random patterns.

A.2. Cap setting results

If the tax aversion effect manifests in self-commitment to withhold effort, we should see more subjects setting low caps that would be likely to bind on their productivity in the tax treatments than in the *Wage Cut Treatment*. (Fig. A2 suggests this is the case, as the CDFs of the caps set in the *Experimenter Tax Treatment* and the *Government Tax Treatment* are above the CDF of caps set in the *Wage Cut Treatment*.) How should we define a low cap in conducting this analysis? Here we are interested in people restricting their own productivity by choosing a cap that is below what they are able to produce in the second 10-minute block. We report on two definitions of a low cap that satisfy that criteria.[22] Under both definitions, decreasing net wage with a tax rather than a wage cut doubles the probability that subjects set a low cap.

---

[22] For robustness, we perform the same analysis with many other definitions of a low cap and always get the same pattern of results.

Table A4
Effect of the net wage change and the tax frame on productivity.

	Difference in productivity (strings completed in second block – strings completed in first block)			Strings completed in second block		
	(1)	(2)	(3)	(4)	(5)	(6)
Net wage of $0.01	−10.01 (2.644)***	−4.030 (3.501)	−4.045 (3.510)	−10.16 (2.595)***	−4.073 (3.417)	−4.074 (3.427)
Tax		−8.295 (3.886)**	−8.756 (4.296)**		−8.446 (3.811)**	−8.472 (4.224)**
U.S. Treasury			0.953 (5.020)			0.0526 (4.973)
Strings in first block				1.164 (0.062)***	1.167 (0.060)***	1.167 (0.060)***
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	325	325	325	325	325	325
R-squared	0.15	0.16	0.16	0.58	0.58	0.58

Table replicates Table 1 from the main text but includes demographic controls. It shows OLS regressions reporting the difference in productivity between the first and second block — regressions (1) through (3) — and the productivity in the second block — regressions (4) through (6). The constant shows the value for the *$0.02 Treatment*. *Net wage of $0.01* indicates that the subject was paid a net wage of $0.01 rather than $0.02. *Tax* indicates that the net wage of $0.01 was implemented using a 50% tax. *U.S. Treasury* indicates that revenue from the tax imposed was sent to the U.S. Treasury rather than paid back to the experimenter. *Strings in first block* is the number of strings the subject completed in the first ten-minute block. Demographic controls include: dummies for gender (i.e., male or female), dummies for year of birth, dummies for current student status (i.e., full time, part time, not enrolled), dummies for monthly household income (i.e., 11 categories from "None" to "$10,000 +", and an option for "don't know / prefer not to answer"), dummies for liberal vs. conservative (i.e., 7 categories from "very conservative" to "very liberal"), and dummies for political affiliation (i.e., Democrat, Republican, Independent, or other). For every demographic variable we include a dummy for non-response to avoid losing data (13 subjects have one or more variables coded as non-response). A constant is no longer reported because it is uninformative given the controls for each demographic group. Robust standard errors are in parentheses.
* Significance at 0.1.
** Significance at 0.05.
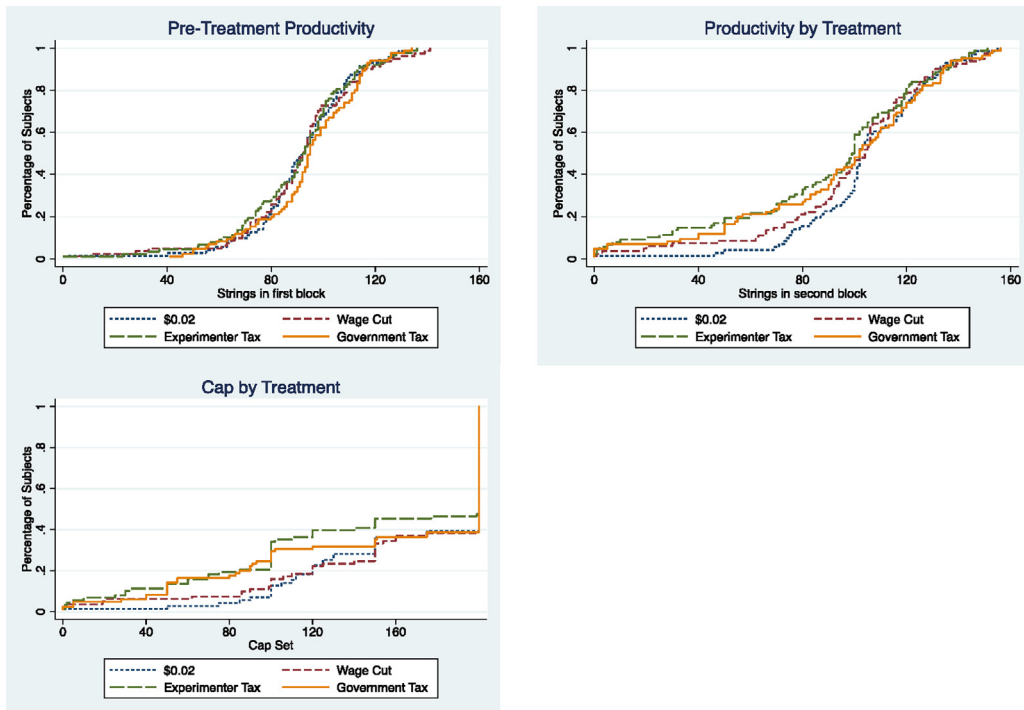*** Significance at 0.01.

Fig. A2. Number of strings and cap by treatment. Figure shows CDFs by treatment of pre-treatment productivity (i.e., number of strings completed in the first block of 10 minutes), productivity (i.e., number of strings completed in the second block of 10 minutes), and cap (i.e., cap set for the second block of 10 minutes).

Table A5
Probability of setting a low cap.

	Set a low cap (low cap defined as $cap < n_1$)		Set a low cap (low cap defined as $cap < 110$, the 25th percentile of all caps)	
	(1)	(2)	(3)	(4)
Net wage of $0.01	0.0177 (0.0401)	0.0171 (0.0398)	0.0320 (0.0592)	0.0283 (0.0548)
Tax	0.0820 (0.0403)**	0.0830 (0.0402)**	0.157 (0.0554)***	0.163 (0.0508)***
Strings in first block		−0.00118 (0.000902)		−0.00713 (0.000851)***
Constant	0.0563 (0.0275)**	0.164 (0.0908)*	0.141 (0.0415)***	0.792 (0.0915)***
Observations	325	325	325	325
R-squared	0.02	0.03	0.04	0.17

Table shows Linear Probability Model (OLS) regressions reporting whether a low cap was set by two different definitions of a low cap. The constant shows the value for the *$0.02 Treatment*. *Net wage of $0.01* indicates that the subject was paid a net wage of $0.01 rather than $0.02. *Tax* indicates that the net wage of $0.01 was implemented using a 50% tax (either to the experimenter or to the U.S. Treasury). A significant effect on *Tax* indicates a tax aversion effect that subjects are more likely to set a low cap by either definition when the wage cut is framed as a tax. *Strings in first block* is the number of strings the subject completed in the first ten-minute block. Robust standard errors are in parentheses.

* Significance at 0.1.
** Significance at 0.05.
*** Significance at 0.01.

First, we define a low cap as being below the number of strings that a subject completed in the first block of 10 minutes. Since subjects are aware the task will stay the same, they should presumably be able to complete approximately as many strings as they did in the previous 10-minute block.[23] Across all subjects, 11.38% set a cap below their previous productivity and all subjects who set a low cap by this definition reach their cap.[24] Using this definition, subjects are twice as likely to set a low cap in the tax treatments than in the *Wage Cut Treatment*. In the *Wage Cut Treatment*, 7.41% of subjects set a low cap; in contrast, 15.61% of subjects in the tax treatments set a low cap.[25] This difference is statistically significant ($p < 0.05$, see Appendix Table A5).[26]

Second, we define a low cap by picking a threshold from the distribution of caps set by all subjects in the experiment. The 25th percentile cap across all 325 subjects in the study was 110, and 77.8% of the 81 subjects in who set a cap below 110 reach it.[27] If we consider a cap low if it falls below the 25th percentile cap, 17.28% of subjects set a low cap in the *Wage Cut Treatment* compared to 32.95% of subjects in the tax treatments.[28] Again this difference is statistically significant ($p < 0.01$, see Appendix Table A5).[29]

More subjects setting a low cap in the tax treatments mirrors the negative effect on time spent working found in Table 2. In addition to responding to the tax treatments on the extensive margin by working less, subjects are using the cap to commit themselves to work less.

Appendix B. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2015.09.022.

References

Abeler, J., Jager, S., 2015. Complex tax incentives. Am. Econ. J. Econ. Policy 7 (3), 1–28.
Alm, J., Torgler, B., 2006. Culture differences and tax morale in the United States and in Europe. J. Econ. Psychol. 27, 224–246.
Alm, J., Sanchez, I., DeJuan, A., 1999. Changing the social norm of tax compliance by voting. KYKLOS 52, 141–171.
Ariely, D., Kamenica, E., Prelec, D., 2008. Man's search for meaning: the case of Legos. J. Econ. Behav. Org. 67, 671–677.
Blumkin, T., Ruffle, B., Ganun, Y., 2012. Are income and consumption taxes ever really equivalent? Evidence from a real-effort experiment with real goods. Eur. Econ. Rev. 56, 1200–1219.
Buell, R., Norton, M., 2011. The labor illusion: how operational transparency increases perceived value. Manage. Sci. 57, 1564–1579.
Buell, R., Norton, M., 2013. Surfacing the Submerged State with Operational Transparency in Government Services. Harvard Business School Working Paper 14-034.
Cabral, M., Hoxby, C., 2012. The Hated Property Tax: Salience, Tax Rates, and Tax Revolts. NBER Working Paper No. 18514.

[23] Of subjects in the *$0.02 Treatment* — who experienced the same wage for the first and second 10-minute block — all who did not set a low cap were able to reach or outperform their productivity from the first ten-minute block.

[24] Meanwhile, only 11.5% of subjects who do not set a low cap by this definition reach their caps.

[25] By this definition, 15.91% of subjects in the *Experimenter Tax Treatment* and 15.29% of subjects in the *Government Tax Treatment* set low caps.

[26] We also get a statistically significant tax aversion effect looking at the average cap set by subjects, but caps can vary dramatically for subjects without changing the probability that they will bind (e.g., from 0 to 50, when it is almost certain to bind, or from 150 to 200, when its almost certain not to bind), so looking at the probability of setting a cap low enough to bind is more relevant.

[27] In contrast, only 3% of subjects who set a cap at 110 or above reach it.

[28] By this definition, 35.23% of subjects in the *Experimenter Tax Treatment* and 30.59% of subjects in the *Government Tax Treatment* set low caps.

[29] We obtain similar results for other thresholds including the 10th percentile threshold cap of 55 strings, the 20th percentile of 100 strings, and the 30th percentile of 125 strings with significance always $p < 0.05$.

Chetty, R., 2009. Sufficient statistics for welfare analysis: a bridge between structural and reduced-form methods. Ann. Rev. Econ. 1, 451–488.
Chetty, R., Looney, A., Kroft, K., 2009. Salience and taxation: theory and evidence. Am. Econ. Rev. 99, 1145–1177.
Doerrenberg, P., Duncan, D., 2014. Experimental evidence on the relationship between tax evasion opportunities and labor supply. Eur. Econ. Rev. 68, 48–70.
Feldman, N., Katuscak, P., 2006. Should the Average Tax Rate be Marginalized? CERGE-EI Working Papers No. 304.
Finkelstein, A., 2009. *E-ZTAX*: tax salience and tax rates. Q. J. Econ. 124, 969–1010.
Fochmann, M., Weimann, J., 2013. The effects of tax salience and tax experience on individual work effects in a framed field experiment. Public Finance Anal. 69, 511–542.
Fochmann, M., Weimann, J., Blaufus, K., Hundsdoerfer, J., Kiesewetter, D., 2013. Net wage illusion in a real-effort experiment. Scand. J. Econ. 115, 476–484.
Hardisty, D., Johnson, E., Weber, E., 2010. A dirty word or a dirty world? Attribute framing, political affiliation, and query theory. Psychol. Sci. 21, 86–92.
Hayashi, A., Nakamura, B., Gamage, D., 2013. Experimental evidence of tax salience and the labor-leisure decision: anchoring, tax aversion, or complexity? Public Finance Rev. 41, 203–226.
Jones, D., 2012. Inertia and overwithholding: explaining the prevalence of income tax refunds. Am. Econ. J. Econ. Policy 4, 158–185.
Kahneman, D., Krueger, A., Schkade, D., Schwarz, D., Stone, A., 2006. Would you be happier if you were richer? A focusing illusion. Science 312, 1908–1910.
Kube, S., Maréchal, M.A., Puppe, C., 2013. Do wage cuts damage work morale? Evidence from a natural field experiment. J. Eur. Econ. Assoc. 11 (4), 853–887.
Kuziemko, I., Norton, M., Saez, E., Stantcheva, S., 2015. How elastic are preferences for redistribution? Evidence from randomized survey experiments. Am. Econ. Rev. 105, 1478–1508.
Lamberton, C., 2013. A spoonful of choice: how allocation increases satisfaction with tax payments. J. Public Policy Mark. 32 (2), 223–238.
Lamberton, C., De Neve, J., Norton, M., 2014. Eliciting Taxpayer Preferences Increases Tax Compliance. Harvard Business School Working Paper 14-106.
Lehmann, E., Marical, F., Rioux, L., 2013. Labor earnings respond differently to income-tax and to payroll-tax reforms. J. Public Econ. 99, 66–84.
Liebman, J., Zeckhauser, R., 2004. Schmeduling. Working Paper.
McCaffery, E., Baron, J., 2006. Thinking about tax. Psychol. Public Policy Law 12, 106–135.
Mettler, S., 2011. The Submerged State. The University of Chicago Press, Chicago, IL.
Mirrlees, J., 1971. An exploration in the theory of optimum income taxation. Rev. Econ. Stud. 38, 175–208.
Muysken, J., Van Veen, T., de Regt, E., 1999. Does a shift in the tax burden create employment? Appl. Econ. 31, 1195–1205.
Reimers, S., 2009. A paycheck half-empty or half-full? Framing, fairness and progressive taxation. Judgm. Decis. Mak. 4, 461–466.
Saez, E., 2001. Using elasticities to derive optimal income tax rates. Rev. Econ. Stud. 68, 205–229.
Sausgruber, R., Tyran, J., 2005. Testing the Mill hypothesis of fiscal illusion. Public Choice 122, 39–68.
Sillamaa, S.A., 1998. How work effort responds to wage taxation: a non-linear versus a linear tax experiment. J. Econ. Behav. Org. 39, 219–233.
Sillamaa, S.A., 1999a. Taxpayer behavior in response to taxation: comment and new experimental evidence. J. Account. Public Policy 18, 165–177.
Sillamaa, S.A., 1999b. How work effort responds to wage taxation: an experimental test of a zero top marginal tax rate. J. Public Econ. 73, 125–134.
Sussman, A., Olivola, C., 2011. Axe the tax: taxes are disliked more than equivalent costs. J. Mark. Res. 48, S91–S101.
Svallfors, S., 2012. Welfare Attitudes in Europe: Topline Results from Round 4 of the European Social Survey. Centre for Comparative Social Surveys, London.
Swenson, C., 1988. Taxpayer behavior in response to taxation: an experimental analysis. J. Account. Public Policy 7, 1–28.
Weber, M., Schram, A., 2013. The Non-Equivalence of Labor Market Taxes: A Real-Effort Experiment. Tinbergen Institute Discussion Paper 13-030/I.