

The Non-Equivalence of Labor Market Taxes: A Real-Effort Experiment*

Matthias Weber[†]

Arthur Schram[†]

December 29, 2014

Abstract

Under full rationality, a labor market tax levied on employers and a corresponding income tax levied on employees are equivalent. With boundedly rational agents, this equivalence is no longer obvious and the different reactions to these two taxes become important for policy making, political economics, and optimal taxation theory. In a real effort laboratory experiment, we study the differential effects of the two taxes on preferences concerning the size of the public sector, subjective well-being, labor supply, and on-the-job performance. Our findings suggest that employer-side taxes induce preferences for a larger public sector. In addition, subjective well-being is higher when the taxes are levied on employers while labor supply is lower, at least at the extensive margin. We discuss three mechanisms that may underlie these results.

JEL classification: C91, H22, H30

Keywords: tax perception; liability side equivalence; political economy; labor supply

*Financial support by The Netherlands' Organisation for Scientific Research (NWO grant 406-11-022) and the Research Priority Area Behavioral Economics of the University of Amsterdam is gratefully acknowledged. Thanks for suggestions and comments go to Ghazala Azmat, Maximilian Hoyer, Boris van Leeuwen, Pedro Robalo, Joep Sonnemans, Johannes Vatter, Jeroen van de Ven, participants of the ESA world meetings in New York, the CREED-CEDEX-CBESS meeting in Norwich, the M-BEES symposium in Maastricht, the workshop The Causes and Consequences of Happiness in Rotterdam, and participants of seminars in Amsterdam and Barcelona.

[†]CREED, University of Amsterdam, and Tinbergen Institute

1 Introduction

Traditional public finance assumes full rationality when analyzing the economic impact of taxes. Under this assumption classic results on tax incidence can straightforwardly be derived, such as liability-side equivalence (LSE). In the words of Joseph Stiglitz:

“It makes no difference whether a tax is imposed on the suppliers of a factor or commodity rather than on the consumers. (...) Taxes induce changes in relative prices, and it is this market response that determines who bears the tax.” (Stiglitz, 2000, p. 514)

Full rationality is a questionable assumption if it aims to describe human behavior in the real world, however. Since at least Simon (1955) the evidence of bounded rationality in economic decisions has accumulated (Conlisk, 1996). For the study of tax incidence, bounded rationality introduces the relevance of issues like tax perception, framing, myopia, or time inconsistency (Bassi, 2010). Assuming full rationality may therefore have far-reaching consequences. Consider, for example, Stiglitz’ assessment. The underlying assumption in the assertion that relative prices determine tax incidence via market responses is that individuals correctly perceive taxes and respond to them in a utility-maximizing manner. If bounded rationality affects either perception or response, prices no longer fulfill this role and LSE is no longer obvious.¹

In this paper, we study the perception and the behavioral responses to distinct labor taxes that are equivalent in the traditional sense. This is important since labor taxes play a major role in all modern economies and many tax policies are still based on the lessons obtained in traditional public finance. The recently emerged field dubbed ‘behavioral public finance’ (e.g. McCaffery and Slemrod, 2006, Mullainathan et al., 2012) intends – among other things – to mend this lack of an empirically sound basis for economic policies. This is the field to which our paper hopes to contribute.² We test one

¹Another assumption commonly made in the traditional public finance literature is that individuals have self-regarding preferences. Numerous papers in behavioral and experimental economics have shown the prominence of other-regarding preferences, however (for a survey, see Cooper and Kagel 2009). If other-regardedness takes processes into account (as opposed to being outcome-based), distinct taxes may not be valued equally. Moreover, if bounded rationality affects the perception of or the response to taxes, other-regarding preferences may inflate the differences.

²A small part of the traditional public finance literature allows for failure of LSE in the labor market due to market frictions. We are not interested here in studying labor market frictions, but in perceptions of and reactions to taxes that are not necessarily rational. Therefore, we study these taxes in a setting where they are by design equivalent under full-rationality, i.e. in a setting without frictions. Rejecting rational reaction to taxes in such a setting also raises doubts about the underpinnings of models based on full rationality combined with market frictions. We will not mention this equivalence every time we compare the taxes – when we mention these taxes, we refer to a tax on the supply side and on the demand side of the labor market that are fully equivalent.

of the prerequisites for tax equivalence to hold. Labor tax equivalence follows from a rational perception of distinct taxes in combination with utility maximizing choices and market forces. We specifically study the first element, i.e., whether the framing of otherwise equivalent taxes affects behavior. Rational perception (which in our case is the absence of a framing effect) is a necessary, but not sufficient condition for tax equivalence to hold. If framing effects are observed, this provides direct evidence against tax equivalence. While most economists think of tax equivalence only in terms of market prices and quantities, there are more ways in which taxes can be equivalent under full rationality. With boundedly rational agents, this equivalence could for example be violated if the distinct taxes induce individuals to prefer the provision of different quantities of a tax financed public good or if the taxes lead to different levels of subjective well-being.

Two ways of taxing labor prevail around the world. One is an income tax levied on employees, the other a payroll tax levied on employers.³ Note that these two types of taxes exist side by side in many countries (often in the form of contributions to social security on both the employee's and the employer's side). This is somewhat surprising from a full rationality point of view (at least, in the absence of market frictions), because under equivalence one would expect the tax to be chosen that minimizes collection and compliance costs. The reason for the co-existence is possibly that people perceive the two taxes differently and react to them in different ways. This is what our paper investigates.

There are many ways in which such taxes may differentially affect people. Here, we list three. First, there may be strong effects on individual political preferences. If perceptions vary across taxes of how much of the public sector is financed by distinct groups in society, opinions on the preferred size of the public sector (or welfare state) are likely to vary as well. This could directly influence voting decisions. Intuitively, it might explain why right-wing politicians tend to favor duties levied on the employees' side while left-wing politicians tend to favor duties levied on the employers' side. Second, many economists are interested in the effects of policies on some index of well-being (representing individual utility or its aggregate, social welfare). Subjective well-being is an obvious first measure of the consequences of policies, including taxes.⁴ People

³Employer payroll taxes often take the form of contributions (for example to social security or health care). Legally, there is a difference between taxes and contributions as in the latter case, employees usually receive an entitlement that they do not receive with a tax. We treat the terms as equivalent here. We use the term 'income tax' for a tax (or contribution) on the employees' (supply) side of the labor market and the term 'employer payroll tax' for a tax (or contribution) on the employers' (demand) side.

⁴For an overview of the literature on the measurement of subjective well-being, see Kahneman and Krueger (2006). For discussions on using such measures for welfare comparisons, see Anand and Van Hees (2006), Schokkaert (2007), and Ferrer-i Carbonell (2013).

simply might be happier under some tax regimes than under others. It is possible, for example, that people get frustrated if they observe a large part of their gross wage being taxed away. In the end, a third effect of labor market taxes is probably to many economists the most obvious. This is their effect on labor supply decisions and on job performance (or total output produced). Individuals may decide to work more or less under one tax regime than under another, either at the intensive margin (hours worked) or at the extensive margin (labor market participation). It may occur, for example, that high gross wages induce people to accept jobs that they would not accept after careful consideration of post-tax income.

We consider the effects of differential perception of theoretically equivalent labor market taxes on each of these three dimensions: political-economic preferences, subjective well-being, and labor supply (together with performance). Simultaneously considering multiple dimensions is important because even simple policy changes – such as the change in the liability side of a labor tax (even with enforced equilibrium wages) – can affect welfare in many ways. It is far from trivial to derive policy implications from these complex effects, but ignoring important dimensions when judging the welfare properties of a policy change can lead to incorrect policy recommendations and possibly very costly mistakes. Our three dimensions represent the three main categories through which individuals may be affected: preferences, well-being, and economic decisions.

For our purpose, observational field data are usually ill-suited, because it is generally difficult to disentangle the numerous effects stemming from broad tax reforms. It is also often impossible to filter out the causes of observed effects (e.g., differences could stem from market frictions or from differences in perception) and moreover, counterfactuals are missing in such data. In addition, field experiments on taxes are almost impossible to implement as governments are highly unlikely to agree to implement a treatment design including proper controls, because not all citizens would be treated equally.⁵

This leaves laboratory experiments as a natural choice to investigate the questions at hand. Even if other empirical methods were feasible, for various reasons such experiments would still be a preferred way to investigate this issue. For one thing, the laboratory allows one to provide a setting that is most favorable for liability side equivalence to hold. All tasks, payoffs, and taxes are more salient and more directly related to decisions than is typically the case outside of the laboratory. Furthermore, institutional frictions are absent and laboratory control allows one to make the taxes equivalent by

⁵Nevertheless, there are a few examples of field experiments on taxation that have been successfully implemented (e.g., the New-Jersey/Pennsylvania Negative Income Tax experiments; see Robins, 1985).

design instead of being equivalent only in general equilibrium. As a consequence, a lack of LSE in the laboratory – where it is given its best shot – would raise serious doubts about its validity outside of the laboratory. In addition, in a careful experimental design one can systematically vary the environment in which the taxes are implemented, which allows one to test the sensitivity of LSE to such changes. For example, we will distinguish between an environment in which proceeds are lost, and one where tax revenues are used to produce a public good. Finally, the laboratory provides the opportunity to directly measure the effects of taxation. In particular, it allows us to directly measure subjects' preferences for the size of the public sector, their subjective well-being and their labor supply responses.

In sum, we examine in a laboratory experiment with human subjects and monetary incentives whether people react differently to an incentive scheme depicting an income tax than to one reflecting a payroll tax levied on employers. By design, both duties are absolutely equivalent under full rationality. To increase the external validity of our laboratory environment, the experiment will require real effort by subjects to earn an income (that may subsequently be taxed). As mentioned above, we will distinguish between two environments. In one, tax revenue does not benefit employers or employees in any way; in the other, proceeds are used to produce a public good. One could argue that the real world is an intermediate case – taxes yield something in return, but the returns perceived by any single tax payer are not always as obvious and direct as is the case of a public good supplied to a small group of subjects in the laboratory. Distinguishing between the two extreme cases allows us to isolate the effects on LSE of (perceived) returns from taxation.

This paper is primarily empirical in the sense that it carefully establishes in a controlled environment whether framing effects exist that contradict LSE. In addition, however, we distinguish between three mechanisms – gross wage illusion, tax aversion, and warm glow – that would explain such framing effects and discuss how these mechanisms interact to predict the experimental results we observe.

Our results show how differences in the way the two taxes are perceived affect behavior in each of the dimensions that we distinguish between. More specifically, employer-side taxes lead to (1) workers preferring a larger public sector; (2) higher subjective well-being of the workers; and (3) lower labor supply (at least at the extensive margin); all in comparison to the case where the taxes are levied on workers' gross income. Each of these effects indicates that it matters who actually transfers the taxes to the government.

This paper intends to add to the literature in the following ways. It is the first to investigate how levying a labor tax on either the employers' or the employees' side of the labor market influences individuals' preferences concerning the size of the public sector. To elicit these preferences in the laboratory, we introduce a novel, incentive compatible approach. This paper is also the first to investigate the effects of the liability side of a labor tax on subjective well-being. In addition, it provides further evidence on the effects of the liability side of a labor tax on labor supply at the extensive and intensive margin and on job performance, being the first to do so in an environment that mimics an employer-employee relationship.

The organization of this paper is as follows. Section 2 reviews the related literature and Section 3 presents the experimental design and procedures. This is followed by Section 4, which contains the hypotheses to be tested and the psychological mechanisms we expect to be at work. Section 5 presents and discusses the results. In Section 6 we discuss some policy implications of these results. Finally, Section 7 concludes.

2 Related Literature

To our knowledge there are no previous studies investigating the effects of the liability side of a labor tax on political-economic preferences (such as preferences for the size of the public sector). Neither do we know of any study of the effects of such taxes on subjective well-being.

There are, however, other studies examining labor supply or job performance under such taxes. A part of this literature is theoretical.⁶ In this short overview we focus on the related empirical, especially experimental, literature.

There is not much non-experimental empirical research that is closely related. A notable exception is Lehmann et al. (2013) who investigate how gross earnings change when income tax rates or payroll tax rates change using recent French data. They find that gross labor earnings respond to changes in the marginal income tax rate while they do not respond to changes in the payroll tax rate, thus rejecting LSE. The authors suggest

⁶There are different approaches to modeling the labor market and thus the impact of labor market taxes. Most prominent are the competitive labor market approach (see Atkinson and Stiglitz, 1980), the efficiency wage theory (Shapiro and Stiglitz, 1984), search and matching models (see Pissarides, 2000) and union bargaining models (see Oswald, 1985). Most of this literature does not allow for liability side non-equivalence. Such non-equivalence can arise in exceptional cases via market frictions, however (e.g. Koskela and Schöb, 1999, Picard and Toulemonde, 2001, Rasmussen, 1997). In general, none of these approaches allow for any non-equivalence due to tax perception. In the presence of labor market frictions, it seems likely that non-rational perception of taxes can considerably amplify or dampen the effects of these frictions.

that this might be due to differential effects of these tax changes on labor supply. Using data from the Netherlands Muysken et al. (1999) present evidence that a larger part of taxes is shifted if they are levied on the employees' side rather than on the employers' side. Holm et al. (1994) use Finnish data in an empirical application of a monopoly union wage determination model and find that increasing the payroll tax rate has a negative effect while increasing the income tax rate has a positive effect on wages. Using Greek data from the early nineteen-nineties, Saez et al. (2012) find that upper income earners do not respond to increases in payroll taxes concerning their labor supply decisions, neither on the intensive nor on the extensive margin.

There is a limited experimental literature on the effects of taxes on labor supply. We know of no such study implementing an employer-employee relationship in the laboratory. Hayashi et al. (2012) experimentally study distinct income tax schemes (specifically, no tax, a flat tax, a progressive tax and a wage subsidy) that are equivalent under full rationality. Their results show that labor supply differs across treatments. Their most robust finding is that subjects choose to supply less labor in the wage subsidy treatment than in the others. Fochmann et al. (2010b) investigate whether the gross wage has an influence on the labor supply decisions of the participants. They find that participants choose to work longer and harder when their gross wage is higher (holding net wages constant). They refer to their finding as 'gross wage illusion'.⁷ Finally, in a field experiment that does not involve taxes, Hossain and List (2012) show that workers in a Chinese factory respond differentially to distinct framing of productivity bonuses. All of these results hint at possible effects that tax framing may have in the labor market.

Other papers examine liability side equivalence of taxes in situations resembling a more general buyer-seller environment. Sausgruber and Tyran (2005) study the perception and effects of direct and indirect taxes. They find that the tax burden associated with an indirect tax is underestimated, which is not the case with a corresponding direct tax. Their study also shows that this can lead to voting for inefficiently high redistribution. Experience seems to weaken this effect, however. Sausgruber and Tyran (2011) add to their previous research by showing that while experience is an effective de-biasing mechanism, pre-vote deliberation about tax regimes is not. Riedl and Tyran (2005) investigate gift exchange markets. Their results support LSE. Finally, Cox et al. (2012) examine tax incidence in double auction and posted offer markets and find that LSE does not hold. All in all, the experimental work on LSE in various environments provides mixed results.

⁷Fochmann and Weimann (2011) elaborate on Fochmann et al. (2010b) and explain gross wage illusion by tax salience.

Considering tax perception more generally, many studies report irrational behavior by laboratory participants. An excellent survey is presented in Fochmann et al. (2010a). An example of this literature is De Bartolome (1995), who shows that many people mistakenly use the average tax rate instead of the marginal tax rate when making investment decisions. Fochmann et al. (2012) study how investment decisions change with the framing of taxes. Their experimental results show that the possibility to deduct losses from an income tax leads to significantly riskier investments (again, their treatments are equivalent under full rationality). Blumkin et al. (2012) find that experimental subjects in the laboratory supply less labor under an income tax than under a consumption tax (when both taxes are equivalent under full rationality). Ullmann and Watrin (2008) conduct experiments showing that people are more likely to evade taxes in a consumption tax environment than in an income tax environment. Such ‘irrationality’ carries over to the field. Chetty et al. (2009) report on a field experiment suggesting that consumers react differently when sales taxes are already included in the price tag than when they are not included in it. The authors also deserve credit for making the concept of tax salience prominent. Finkelstein (2009) provides evidence that tolls become less salient when collected electronically and that drivers’ behavior then becomes less elastic to the level of the toll.

3 Experiment

The experiment was conducted at the CREED laboratory at the University of Amsterdam in February 2012 with a total of 240 subjects recruited from the CREED subject pool. Participants were mainly undergraduate students, slightly less than half were female and roughly 60% majored in economics or business. The experiment was programmed in PHP/MySQL. Every participant received a show up fee of 7 euros. During the experiment, ‘points’ were used as currency. These were exchanged into euros at the end of each session at an exchange rate of 1 euro per 600 points. The experiment lasted between 90 minutes and 2 hours in total and participants earned on average about 22 euros, including the show up fee. Before starting, the participants had to answer control questions to make sure that they understood the instructions. The experiment did not start until all participants had successfully answered these questions. Appendix A provides a transcript of the instructions and test questions, Appendix B contains screenshots. During the experiment, subjects received no information on the choices or the performance of other subjects.

Twelve sessions were run, three each for four distinct treatments.⁸ When scheduling, we distributed the sessions in a balanced way over mornings and afternoons and across the different days of the week.

3.1 Treatments

The design is a 2×2 factorial, between-subject design. Subjects are either employer or employee. They are allocated to groups consisting of one employer and five employees. Those in the role of employees work on a task for which they receive performance-based remuneration; the employer receives earnings depending on the performance of the employees in the same group. The form this incentive scheme takes is one of the treatment variables. In one case, employees receive a gross wage, from which a duty is subsequently deducted as a tax. In the other, employers pay the duty and employees receive a (lower) net wage. Note that this corresponds naturally to labor market taxes levied either on the employees' or on the employers' side. What happens with the tax proceeds is varied in the second treatment variable. The tax proceeds are either taken away ('nothing in return') or used to produce a public good (which is called 'common fund' in the experiment). Table 1 summarizes the design. The acronyms *EN*, *IN*, *EP* and *IP* for the four distinct treatments will be used regularly in the remainder. In parentheses are the numbers of subjects per treatment. Figure 1 shows a sketch of earnings and taxes.

Table 1: 2×2 design

	Employer payroll tax	Income tax
Nothing in return	<i>EN</i> (60)	<i>IN</i> (60)
Public good	<i>EP</i> (60)	<i>IP</i> (60)

Notes: Cells indicate the acronyms used for the treatment. Each treatment combines a tax levied either on the employers' (*E*) or employees' (*I*) side with the case where tax proceeds are either lost (*N*) or used to produce a public good (*P*). In parentheses are the numbers of subjects per treatment.

The tax rate used is 40% in the income tax treatments and 66.7% of the corresponding lower wage in the employer payroll tax treatments. This ensures equivalence between the income tax and the employer payroll tax in terms of employees' net earnings and

⁸A pilot was run in the summer of 2011, as documented in Weber (2011). The new sessions differ significantly from the pilot; the most important changes are the introduction of a mechanism to measure preferences for the size of the public sector and the introduction of a leisure task. More information is available upon request.

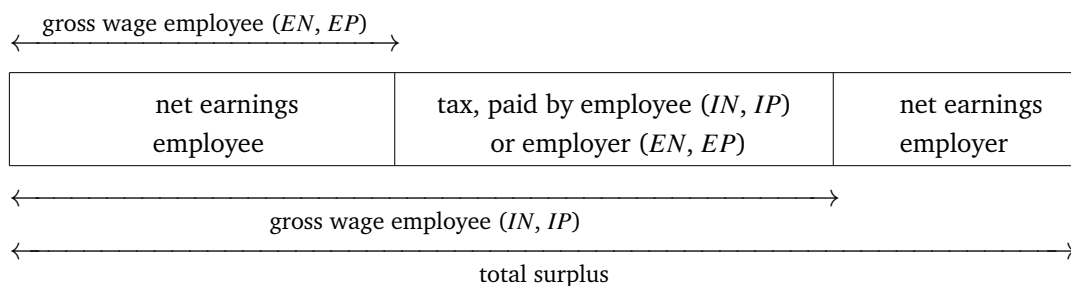


Figure 1: Wages and taxes in the experiment

employers' gross costs, both for the nothing in return and the public good treatments.⁹ Hence, differences between EN and IN and between EP and IP reflect only framing effects. The public good in the relevant treatments is produced with a multiplication factor of 1.3 (meaning that the tax revenue allocated to the common fund is increased by 30%) and its returns are equally distributed among all employees in a group at the end of the experiment.¹⁰ The wages were chosen such, that the net wage in a nothing-in-return treatment is equal to the net wage plus the return from the public good from one's own tax payment in the public good treatment.¹¹

3.2 Course of Events

At the beginning of each session, subjects are randomly divided into groups of six. One subject in each group is randomly determined to be 'employer', the other five are 'employees'. The group composition remains fixed throughout the experiment. The experiment consists of multiple parts. The participants receive the instructions for a

⁹Assume that the employer's labor costs are 1 euro. In the income tax treatments this is the employee's gross wage and after an income tax of 40%, the employee has 60 cents left. This is equivalent to the case of the payroll tax, where employee receive a wage (gross and net) of 60 cents and the employer pays 67.7% of this wage (i.e. 40 cents) as a tax on top of the wage, making her labor costs again equal to 1 euro.

¹⁰As a consequence, a tax revenue of x points in a group with 5 employees, yields $\frac{1.3x}{5}$ points from the public good for each employee. The public good is not supplied to employers, because we envisage a public good related to income security, e.g., unemployment benefits.

¹¹Recall that in the public good treatments the net earnings consist of the net wage plus the returns from the public good. The returns from the public good can be split in a part that is due to own tax payment and a part that is due to the taxes paid by the other employees in the group. Consider a task an employee faces. Denote by W_P and W_N the net wages for this task in the public good and the nothing in return treatments, and by $\gamma \cdot W_P$ the tax paid, which is a (mandatory) contribution to the public good, where applicable. In the nothing in return treatments the employee receives W_N , which compares to $W_P + \frac{\gamma W_P \cdot 1.3}{5}$ in the public good treatments. W_P and W_N are chosen to equate these two returns. Subjects receive no information on the choices or performances of others during the experiment. Hence, their choices cannot influence others. Note that this means that average total net earnings are higher in the public good treatments than in the treatments without public good, because employees also benefit from returns of the public good that are due to taxes paid by other employees in their group.

part only after the previous parts have been completed. The terms ‘employer’ and ‘employee’ are used intentionally, as is the term ‘wage’. Neutral wording is chosen for the duties, terms such as ‘income tax’ or ‘employer payroll tax’ are not used to avoid (unmeasured) preconceptions that some subjects might have with respect to these terms. Figure 2 gives a schematic overview of the timeline; the different parts will be explained in the following.

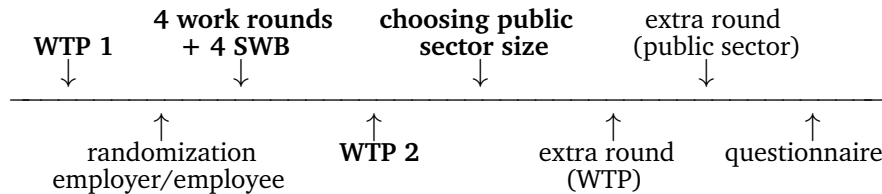


Figure 2: Timeline of the experiment

Notes: The parts encompassing decisions or actions leading to dependent variables are in bold. The acronyms depict the elicitation of subjects’ willingness to pay for an extra round (WTP) and subjective well-being (SWB). Instructions are only given right before the respective task, except for the beginning where instructions for WTP 1 and the four regular rounds are (of course) given. Subjects receive information on the randomly drawn price for the first and the parameters for the second extra round right before the respective round starts. Information on (payments due to) other subjects’ decisions is only provided right before the questionnaire.

3.2.1 Real-Effort Task and Leisure Option

The experiment involves a real-effort work task, which is the following. Each employee sees two 10×10 matrices on the screen that are filled with randomly generated two-digit numbers. Figure 3 shows a screenshot from the work task (taken from treatment *IN*; see Appendix B for a larger version). One matrix is shown on the left half of the screen and the other one on the right half. The employees’ job is to find the largest number in the left matrix and the largest number in the right matrix and to add these two numbers up. For the summation, the participants are provided with pocket calculators. After answering, irrespective of whether the answer is correct, a new pair of matrices appears. This means that subjects have only one attempt to provide the correct answer. Each employee faces a maximum of 30 of these problems, which is much more than they can actually solve correctly in one round, which lasts for 8 minutes. This limit and the way the random numbers are generated make guessing a very unsuccessful strategy. Only the number of correct additions matters, there is no punishment for incorrect additions. While the employees are doing this task, they can see at the top of

the monitor the amount they will receive if the next number they enter is correct and, where applicable (i.e., when tax is levied on the employees), the amount that will be deducted from it (as a tax). Furthermore, they can see how much they have already earned and, where applicable, the amount that will be deducted from it. They can also see the number of correct and incorrect additions so far and the remaining time. The total number of correctly solved problems is a measure of job performance. This procedure is repeated in four independent and identical rounds. During these rounds, employers do not need to do anything. Employers are passive in the experiment, but they nevertheless play an important role for two reasons. First, their presence increases the external validity of the framing in terms of wages and tasks. Second, subjects with other-regarding preferences are affected differently when (as in the world outside of the laboratory) their decisions have real consequences for another person.

Amount of correct additions in this round so far:	1	Round 1 of 4	Time: 06:53
Amount of incorrect additions in this round so far:	1		
Remaining problems that can be solved in this round:	28	Total wage earned in this round:	429 (minus 171.6)

If your next answer is correct you will receive 390 points (from which 156 will be deducted).

<table border="0" style="width: 100%; text-align: center;"> <tr><td>38</td><td>39</td><td>24</td><td>30</td><td>58</td><td>67</td><td>53</td><td>60</td><td>53</td><td>58</td></tr> <tr><td>73</td><td>20</td><td>12</td><td>79</td><td>56</td><td>32</td><td>29</td><td>24</td><td>79</td><td>40</td></tr> <tr><td>24</td><td>30</td><td>43</td><td>25</td><td>21</td><td>59</td><td>25</td><td>73</td><td>78</td><td>41</td></tr> <tr><td>67</td><td>23</td><td>10</td><td>77</td><td>35</td><td>30</td><td>15</td><td>25</td><td>40</td><td>76</td></tr> <tr><td>48</td><td>58</td><td>20</td><td>36</td><td>28</td><td>35</td><td>15</td><td>29</td><td>49</td><td>57</td></tr> <tr><td>33</td><td>71</td><td>30</td><td>53</td><td>44</td><td>72</td><td>65</td><td>55</td><td>56</td><td>49</td></tr> <tr><td>73</td><td>62</td><td>39</td><td>50</td><td>78</td><td>35</td><td>72</td><td>22</td><td>69</td><td>44</td></tr> <tr><td>54</td><td>14</td><td>71</td><td>50</td><td>63</td><td>42</td><td>71</td><td>52</td><td>70</td><td>17</td></tr> <tr><td>25</td><td>12</td><td>13</td><td>23</td><td>62</td><td>68</td><td>71</td><td>41</td><td>65</td><td>43</td></tr> <tr><td>42</td><td>16</td><td>24</td><td>74</td><td>38</td><td>68</td><td>32</td><td>56</td><td>65</td><td>74</td></tr> </table>	38	39	24	30	58	67	53	60	53	58	73	20	12	79	56	32	29	24	79	40	24	30	43	25	21	59	25	73	78	41	67	23	10	77	35	30	15	25	40	76	48	58	20	36	28	35	15	29	49	57	33	71	30	53	44	72	65	55	56	49	73	62	39	50	78	35	72	22	69	44	54	14	71	50	63	42	71	52	70	17	25	12	13	23	62	68	71	41	65	43	42	16	24	74	38	68	32	56	65	74	<table border="0" style="width: 100%; text-align: center;"> <tr><td>83</td><td>57</td><td>25</td><td>13</td><td>51</td><td>34</td><td>66</td><td>42</td><td>74</td><td>63</td></tr> <tr><td>12</td><td>69</td><td>29</td><td>28</td><td>19</td><td>37</td><td>53</td><td>77</td><td>57</td><td>74</td></tr> <tr><td>22</td><td>22</td><td>58</td><td>46</td><td>75</td><td>82</td><td>66</td><td>82</td><td>13</td><td>70</td></tr> <tr><td>14</td><td>50</td><td>47</td><td>58</td><td>21</td><td>76</td><td>56</td><td>81</td><td>19</td><td>30</td></tr> <tr><td>27</td><td>62</td><td>52</td><td>38</td><td>59</td><td>36</td><td>54</td><td>42</td><td>68</td><td>74</td></tr> <tr><td>67</td><td>21</td><td>78</td><td>64</td><td>25</td><td>67</td><td>78</td><td>77</td><td>46</td><td>83</td></tr> <tr><td>20</td><td>73</td><td>63</td><td>13</td><td>40</td><td>51</td><td>45</td><td>57</td><td>15</td><td>45</td></tr> <tr><td>74</td><td>83</td><td>81</td><td>52</td><td>46</td><td>71</td><td>43</td><td>34</td><td>42</td><td>37</td></tr> <tr><td>37</td><td>62</td><td>31</td><td>34</td><td>13</td><td>23</td><td>20</td><td>78</td><td>36</td><td>62</td></tr> <tr><td>54</td><td>75</td><td>55</td><td>67</td><td>65</td><td>40</td><td>59</td><td>34</td><td>23</td><td>28</td></tr> </table>	83	57	25	13	51	34	66	42	74	63	12	69	29	28	19	37	53	77	57	74	22	22	58	46	75	82	66	82	13	70	14	50	47	58	21	76	56	81	19	30	27	62	52	38	59	36	54	42	68	74	67	21	78	64	25	67	78	77	46	83	20	73	63	13	40	51	45	57	15	45	74	83	81	52	46	71	43	34	42	37	37	62	31	34	13	23	20	78	36	62	54	75	55	67	65	40	59	34	23	28
38	39	24	30	58	67	53	60	53	58																																																																																																																																																																																																
73	20	12	79	56	32	29	24	79	40																																																																																																																																																																																																
24	30	43	25	21	59	25	73	78	41																																																																																																																																																																																																
67	23	10	77	35	30	15	25	40	76																																																																																																																																																																																																
48	58	20	36	28	35	15	29	49	57																																																																																																																																																																																																
33	71	30	53	44	72	65	55	56	49																																																																																																																																																																																																
73	62	39	50	78	35	72	22	69	44																																																																																																																																																																																																
54	14	71	50	63	42	71	52	70	17																																																																																																																																																																																																
25	12	13	23	62	68	71	41	65	43																																																																																																																																																																																																
42	16	24	74	38	68	32	56	65	74																																																																																																																																																																																																
83	57	25	13	51	34	66	42	74	63																																																																																																																																																																																																
12	69	29	28	19	37	53	77	57	74																																																																																																																																																																																																
22	22	58	46	75	82	66	82	13	70																																																																																																																																																																																																
14	50	47	58	21	76	56	81	19	30																																																																																																																																																																																																
27	62	52	38	59	36	54	42	68	74																																																																																																																																																																																																
67	21	78	64	25	67	78	77	46	83																																																																																																																																																																																																
20	73	63	13	40	51	45	57	15	45																																																																																																																																																																																																
74	83	81	52	46	71	43	34	42	37																																																																																																																																																																																																
37	62	31	34	13	23	20	78	36	62																																																																																																																																																																																																
54	75	55	67	65	40	59	34	23	28																																																																																																																																																																																																

Largest number in the left matrix plus largest number in the right matrix:

Figure 3: Screenshot during a work round (taken in treatment *IN*)

Employers receive a net payment of 49.8 points per correct addition by any of their five employee's. Net payments for the employees are linearly decreasing in the number of attempts (but are restricted to be non-negative). If employees solve the first problem correctly in the nothing in return treatments, they receive 280.8 points. With each attempt (whether correct or not) the payment for the next correct addition decreases by 23.4 points.¹² In the public good treatments, these numbers correspond to the net return from own performance (the direct net wage plus the return from the public good that is due to own performance, for the calculation see Footnote 11).

We provide employees with an outside option. Instead of working, they can also choose a leisure task, which is framed as a 'fixed payment option'. At any moment during the work rounds, employees can click on a button 'Go to fixed payment option'. After doing so, they are shown a largely empty screen for the rest of the round and receive a fixed payment of 2.2 points per second remaining. They cannot return to solving problems in the same round. Note that the total amount of time (in seconds) spent in the 'work-mode' provides a natural measure of labor supply at the intensive margin.

3.2.2 Measuring Subjective Well-Being

After each round, the employees are shown a screen depicting their gross wage and the number of points paid as tax (if applicable) in the preceding round. Then, participants are surveyed to measure their subjective well-being using a self-assessment manikin (the SAM-V-9; Irtel, 2007, Lang, 1985, Bradley and Lang, 1994). This measure of subjective well-being is also referred to in the literature as satisfaction, happiness, or experienced pleasure. Subjects are asked to report how they are feeling by clicking on one of nine images on the manikin. These images are drawings depicting emotions ordered from least happy to most happy, thus yielding a score from 1 (low pleasure) to 9 (high pleasure). The number is referred to as the 'self-assessment score'. We will use the sum of these scores over the four rounds as our measure of subjective well-being. The self-assessment manikin is shown in Figure 4.

¹²Decreasing payments make it more likely that people make use of the fixed payment option (the leisure task) at some point, i.e. they lead to more interior solutions concerning the time spent working. They can be seen as representing diminishing marginal revenue.

Formally, net earnings from correctly solving a problem in the nothing in return treatments can be written as $\pi = \max(280.8 - 23.4x, 0)$, where x is the number of problems the employee has previously attempted to solve in the same round. This is also the gross wage in the treatment *EN* (net wage equals gross wage), while the gross wage in the treatment *IN* is $\pi = \max(468 - 39x, 0)$, which leads, with a tax rate of 40%, to the same net wage as in *EN*.

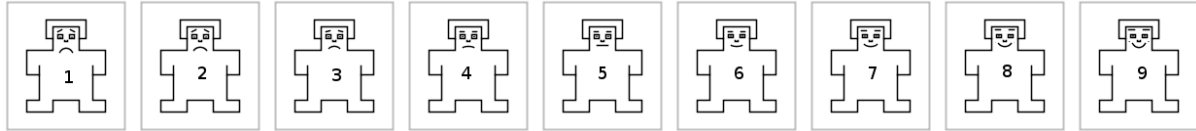


Figure 4: Subjective well-being self-assessment manikin

Notes: After each round, subjects are asked to choose the one of the nine figures that best describes their current emotion.

3.2.3 Measuring Labor Supply at the Extensive Margin

After finishing the instructions for the part comprising the four work rounds, but before being told whether they are employers or employees, subjects are asked to state their willingness to pay for participating in an extra work round after the four regular rounds will have been completed.¹³ For this purpose a BDM mechanism is used (Becker et al., 1964). The price of participation in the extra round is determined randomly (drawn from a uniform distribution between 1056 and 2400 points). The lower limit corresponds to the amount earned after immediately choosing the fixed payment option, the upper limit is a number slightly higher than our expectations of maximum possible earnings in one round. It is randomly determined whether the extra round takes place or not. If it takes place and if the price is lower than the amount stated by an employee, this employee pays the price and works (and gets paid) for another round. If the price is higher than the bid of the employee, the employee neither pays for nor works in another round. Subjects not participating remain seated until all participants have finished. If subjects have a true valuation for participating in this extra round it is a dominant strategy to bid this true valuation. In the treatments with public good the returns from the public good are split among all employees of a group, those working in the extra round and those not working. The stated willingness to pay provides us with a (first) measure of labor supply at the extensive margin.¹⁴

After having finished the regular four work rounds, employees are confronted with the same BDM mechanism again. They are told that the extra round corresponding to the number they enter after having completed the regular rounds will be played out if

¹³Because this is asked before some are selected to be employers, all provide a willingness to pay; for those randomized to be employers the revealed willingness to pay has no further consequence.

¹⁴In reality, people usually do not start participating in the labor market by just clicking on a button somewhere. Rather, participation costs money, time, and effort. One way to think of this is investment in schooling or acquiring specific knowledge in order to be able to participate in the labor market. More generally, one can think of all kinds of costs associated with beginning a new job, including looking for suitable job offers, writing and sending applications, going to interviews, maybe even moving to a different city, etc. Therefore, we consider this willingness to pay a better measure of labor supply at the extensive margin than a mere statement whether one wants to participate or not.

and only if the extra round corresponding to the willingness to pay elicitation before the regular rounds will not be played out. This willingness to pay after the four work rounds provides us with a (second) measure of labor supply at the extensive margin. Our procedures imply that an extra round based on stated willingness to pay always takes place, at a later point (though, whether an individual employee participates in it depends on her stated willingness to pay). It consists of exactly one round, based on either the first (pre-play) BDM mechanism or the second (post-play).

3.2.4 Eliciting Preferences for Public Sector Size

After this second statement of willingness to pay, subjects are told that there will be yet another round. All employees participate in this round, which will take place with new rules. The rules differ from the regular rounds as follows. In the nothing in return treatments (*EN* and *IN*), a public good is introduced, such that the taxes are no longer lost, just as in the public good treatments. Now, the tax rate and the multiplication factor for the public good are no longer given. Instead, they are chosen by the employees in a random dictator style, using the following mechanism. Subjects are presented with a slider, as shown in Figure 5. Each position of the slider represents a unique combination of tax rate and multiplication factor. At the left end of the slider, the tax rate is zero, while the multiplication factor for the public good is high. When moving the slider from left to right, the tax rate increases, and the multiplication factor decreases. The trade-off between tax rate and multiplication factor can be interpreted as a diminishing marginal productivity of the public sector; the higher the tax revenue is, the lower is the efficiency of public good production. While one could also have subjects choose the tax rate for a fixed multiplication factor, we prefer to explicitly allow for this trade-off in order to accommodate preferences for a point at the interior of the slider (with fixed multiplication factor, one would expect many corner solutions where subjects prefer either 0 or maximal tax rates). The slider has 101 different positions, yielding a number between 0 and 100, where 0 corresponds to the leftmost position of the slider, which is used as the default position.¹⁵ The number corresponding to the chosen slider position provides us with a measure of the subject's preferred size of the public sector. After all subjects have chosen a slider position, one employee in each group is randomly selected and her choice is used for this extra round. Note that the employer payroll tax and the

¹⁵The multiplication factor of the public good is 3 at the default position on the left end and 0.75 at the right end. The tax is 0 at the left end and 75% of the employer's labor cost at the right end; the tax is always expressed as a percentage of the employee's (gross) wage. Note that choosing a position where the multiplication factor is less than 1 is dominated in the sense that subjects would always earn more at lower tax/higher productivity rates.

Please choose your most preferred position of the slider.

Multiplication factor of the common fund:	3.00
Contribution to the common fund, expressed as a percentage of your wage:	0.00




Figure 5: Slider for the elicitation of public sector size preference

Notes: The slider is at the default position. When moving the slider from left to right the multiplication factor of the public good (framed as a ‘common fund’) decreases while the contribution to the public good (i.e. the tax rate) increases.

income tax treatments are still absolutely equivalent, whereas the nothing in return and the public good treatments are now somewhat ‘less equivalent’. Subjects in the nothing in return treatments have had no experience with the public good prior to this round. Furthermore, net payments in the nothing in return treatment are not adjusted to the levels used in the public good treatment in order to avoid subjects having to adapt to a new payment schedule for the extra round. As a consequence, payoffs here are slightly higher in the treatments *EN* and *IN* than in *EP* and *IP*.

Finally, the two extra rounds are played out (one originating from the willingness to pay for an extra round and one from the choice of public sector size parameters).¹⁶ The information on the randomly drawn price and the public sector size parameters selected in random dictator style is given to subjects before the respective round starts. After these two rounds information on payments stemming from others’ public goods contributions (i.e., their tax payments) and on the outcomes of the lotteries are revealed. Before being paid, participants are asked to fill out a short questionnaire, including questions concerning gender, age, field of study and experience in laboratory experiments (see Appendix A).

¹⁶Before these rounds are played, we elicit loss aversion using the test developed by Fehr and Goette (2007); see Appendix A. The measure turns out not to lead to additional insights in the data analysis – we attribute this to the facts that we are not dealing with risky choices (for which the test is designed) and to the fact that the measure is quite rough (most subjects end up in the same category). Therefore, this measure will not be discussed further in this paper. More information is available from the authors.

4 Hypotheses and Mechanisms

The main goal of this paper is to test the effects of the liability side of a labor tax, represented by pure framing, on outcomes of the three dimensions we discussed. As discussed above, the major advantage of conducting these tests in the laboratory is that it offers the control to hold everything, except for the policy variable of interest, constant. Possible disadvantages are related to a possible lack of external validity due to, e.g., the artificiality of the setting or the subject pool used. These disadvantages are largely avoided, if one focuses on directions of treatment effects (rather than on levels and quantifications; see Charness and Kuhn, 2011, Falk and Heckman, 2009, or Schram, 2005). This is our approach, here. We derive null-hypotheses about treatment effects from rational choice behavior. The most general two-sided alternative hypotheses can be formulated without testing any particular theory. After stating the hypotheses, we nevertheless discuss psychological mechanisms that we expect may govern participants' behavior and explain how they interact to lead to the alternative hypotheses (including some further predictions that can be made if these mechanisms are at work).

4.1 Hypotheses

We use the predictions of fully rational selfish agents for the null-hypotheses as this is the standard benchmark in economics, but note that the two taxes are also equivalent assuming various other outcome based preferences, such as for example inequity aversion (Fehr and Schmidt, 1999). The incentive schemes we use are equivalent by design, which means that no general equilibrium mechanisms are needed to arrive at the full-rationality outcome. All null-hypotheses reflect no difference in outcomes between the respective employer payroll tax and income tax treatments.¹⁷

Consider the dependent variables derived from subjects' choices as described in Section 3.2. We define variables straightforwardly, denoting by ps the preference concerning the size of the public sector (as measured by the chosen slider position); by sw the subjective well-being (as measured by the four choices with the self-assessment manikin); by $le1$ and $le2$, respectively, the first and second measures of labor supply at the extensive margin (as given by choices in the two BDM tasks); by li the labor supply at the intensive margin (measured by the number of seconds in the work mode over

¹⁷Strictly speaking, there are no rational choice predictions concerning subjective well-being, because decisions are not incentivized. We nevertheless use the null-hypotheses of no difference in outcomes, which is also here the obvious choice.

the four regular rounds); and by jp the job performance (measured by the total number of problems solved correctly). We denote null-hypotheses by $H_{\cdot,0}$ and alternative hypotheses by $H_{\cdot,A}$. All alternative hypotheses used are two-sided.

We write the hypotheses as general hypotheses between the employer payroll tax condition and the income tax condition (denoting this by E and I , respectively). They can then be tested either between the corresponding treatments separately (EN and IN or EP and IP) or for both treatments together in regressions controlling for how the tax proceeds are used. This gives the following hypotheses.¹⁸

First, concerning preference for the size of the public sector, we test

$$H_{ps,0} : ps_E = ps_I \text{ vs } H_{ps,A} : ps_E \neq ps_I.$$

Concerning subjective well-being, we test

$$H_{sw,0} : sw_E = sw_I \text{ vs } H_{sw,A} : sw_E \neq sw_I.$$

For the two measures of labor supply at the extensive margin, we test

$$H_{le1,0} : le1_E = le1_I \text{ vs } H_{le1,A} : le1_E \neq le1_I, \text{ and}$$

$$H_{le2,0} : le2_E = le2_I \text{ vs } H_{le2,A} : le2_E \neq le2_I.$$

Finally, for the decision on labor supply while working (labor supply at the intensive margin) and for job performance, we test

$$H_{li,0} : li_E = li_I \text{ vs } H_{li,A} : li_E \neq li_I, \text{ and}$$

$$H_{jp,0} : jp_E = jp_I \text{ vs } H_{jp,A} : jp_E \neq jp_I.$$

4.2 Mechanisms

Here, we discuss which psychological mechanisms we expect may govern participants' behavior. Note that we allow these mechanisms to have an impact, that differs across the decisions and tasks participants face. While it is not the main goal of this paper to test the mechanisms per se, in some cases our distinction between nothing-in-return

¹⁸As we are mainly interested in the effects of the tax liability side, we do not formulate or test hypotheses between the nothing in return and the public good treatments. While the treatments are all equivalent with selfish and fully rational agents, there can also be differences between a nothing in return and a corresponding public good treatment with boundedly rational agents (for example if a dollar of wage is more salient than a dollar received from the public good due to own performance).

and public good treatments leads to testable predictions of the working of these mechanisms within the design of this experiment.

The three mechanisms that we expect to govern individuals' choices are the following.

- (i) We regard a dollar of wage more salient than a dollar of tax. As a consequence, people will tend to focus more on gross wages than on net wages (as they do not fully take the taxes into account), which is what Fochmann et al. (2010b) call 'gross wage illusion' (as observed by Fochmann et al., 2010b, and Hayashi et al., 2012, in different settings).
- (ii) We expect people to consider a tax payment as a loss that, *ceteris paribus*, they would prefer to avoid. This effect is expected to be asymmetric, in that an employee sees a tax paid by herself as more of a loss than a tax paid by her employer. We use the term 'tax aversion' for this effect.¹⁹
- (iii) Individuals having other-regarding preferences may derive positive utility when a public good is provided to others using tax payments they made ('warm glow', Andreoni, 1990). This warm glow effect here depends on the perception of the tax – it exists if individuals feel more that it is actually their money that is used to benefit others under an income tax than under an employer payroll tax.

We describe now, how gross wage illusion, tax aversion and warm glow may interact and influence participants behavior in the decisions and tasks we are testing (this can be seen as a motivation of the alternative hypotheses above).

We start with preferences regarding the size of the public sector. Gross wage illusion may lead to a higher preferred size in the income tax treatments than in the payroll tax treatments. This is the case if there is a positive income effect for the demand for public goods, i.e., higher (perceived) income yields a preference for a larger public sector (although possible, we expect this effect to be very small if it exists). Tax aversion has the opposite effect, however, because it implies that for any given size of the public sector the perceived tax costs needed to achieve it are smaller when taxes are levied on the employers' side of the labor market. Hence, a larger public sector is preferred in the payroll tax treatments. The warm glow effect goes in the same direction as gross wage illusion. Individuals subject to warm glow may prefer a larger public sector when they are funding it. With two effects pointing in the direction of a higher preferred size in

¹⁹This term has been used in similar but not identical ways in the literature, for example for negative reactions on the mere label 'tax' (McCaffery and Baron, 2006) or for voting against taxation when taxation is in the individual's material self-interest (Kallbekken et al., 2010).

the income tax treatment and one effect predicting the opposite, the aggregate effect could go either way.²⁰

Next, consider subjective well-being. In the nothing in return treatments, gross wage illusion implies that people perceive a higher income in the income tax treatment than in the payroll tax treatments. A natural hypothesis is then that they will feel better in the former case. Tax aversion points in the opposite direction: people will be less happy when they feel that the tax represents money taken away from them. If there is a public good, also warm glow may be present. This effect aligns with the effect of gross wage illusion on subjective well-being. It predicts that people will feel better in the income tax case, because they conceive themselves as helping.

Finally, for labor supply and job performance, we again need to consider the aggregate of the three effects. Gross wage illusion directly implies that people will supply more labor (at both the intensive and extensive margins) and perform better in the income tax treatments.²¹ Tax aversion effects labor supply decisions in the opposite direction. This is because such decisions are assumed to be based on the (subjective) costs and benefits of exerting effort. Taxes on the employees' side add to the perceived costs of exerting effort more than taxes on the employers' side. In the treatments with a public good the warm glow effect will yield higher labor supply and better performance in the income tax treatment because of the perceived extra benefits related to helping others. Once again, gross wage illusion and warm glow point in one direction, and tax aversion points to the opposite.

On top of the fact that we expect these mechanisms to lead to the rejection of rational behavior null-hypotheses, one part of these mechanisms can be tested directly within the framework of this experiment. The warm glow effect when considering subjective well being or labor supply and job performance only exists in the public good treatments. Thus, if our mechanisms drive the results and warm glow plays a role, there should be a positive interaction effect between the tax treatment condition and the public good treatment condition for subjective well-being, labor supply, and job performance. We will test this, below.

²⁰Note that we assume that the framing of the tax has no influence on subjects' beliefs on how others react to the framing.

²¹This is the case if a higher wage leads to higher labor supply, as seems to be generally the case in the real world (Evers et al., 2008) and arguably even more so in laboratory experiments.

5 Results

We have collected data from 10 groups per treatment. Because we only use data obtained from employees to test our hypotheses, this gives 50 observations per treatment, except for the first measure of labor supply at the extensive margin, where we have 60 observations per treatment (as this was measured before the randomization into employees and employers).²² Because subjects receive no feedback about others' decisions until the end of the experiment, we can treat observations across individuals as statistically independent.

Throughout this section, we use censored regression models to analyze our results. We pool the data of the two employer payroll tax treatments and of the two income tax treatments (each with and without provision of a public good). This allows us to observe the effects of the interaction between the tax type and the use of the tax proceeds (i.e., the interaction between a tax type dummy and a public good provision dummy). Except for the two dummies representing the treatment conditions and their interaction, all regressions include the covariates age, gender, field of study (economics and sciences vs. other fields), and experience with laboratory experiments. We present these same specifications for each dependent variable but note that the results are robust to changes, such as taking out a covariate, changing the definition of field of study (e.g. economists vs. others), or using OLS instead of censored regressions. Appendix C.1 presents summary statistics of our main variables and shows the results of non-parametric tests, which draw on very unrestrictive assumptions; these tests yield the same conclusions as presented here.²³ In what follows, we not only test the main hypotheses from Section 4, but also consider the implications of the results for the mechanisms discussed.

5.1 Public Sector Size Preference

Table 2 shows the coefficient estimates for the independent variable ps representing preferred public sector size. Recall that we are especially interested in the tax treatment dummy.

²²Due to computer problems, one observation in IN has missing data on sw and one observation in IP has missing data on sw , li , and jp .

²³Appendix C.1 also presents the data split according to gender. We observe gender effects for some of our results. This could mean that the mechanisms we distinguished between might affect men differently than women. We did not intend to study gender effects and therefore the analysis remains explorative, however.

Table 2: Results for public sector size preference (ps)

	Coefficient	Std. error
Intercept	72.30***	(19.58)
Age	-1.40*	(0.83)
Gender (1=male)	4.18	(4.03)
Studies (1=econ+science)	-2.41	(4.50)
lab experience (1=yes)	9.67	(6.04)
Tax (1=income tax)	-17.15***	(5.54)
PG (1=public good)	-9.91*	(5.46)
Tax * PG	9.04	(7.77)

Notes: The dependent variable in this regression is the preference for public sector size (variable ps). Individual outcomes are integers between 0 and 100 corresponding to the position of the slider, larger numbers represent a preference for a larger public sector. 13 observations are left censored, 176 uncensored, and 11 right censored. *, **, and *** stand for significance at the 10-, 5-, and 1-percent level.

The results show that subjects prefer a significantly larger public sector under an employer payroll tax than under an income tax – the coefficient of the tax condition dummy is negative (and significant at the 1%-level). We thus reject the null-hypotheses $H_{ps,0}$ in favor of the alternative hypothesis. The interaction coefficient is not significant, which is not particularly surprising, given that the mechanisms we outlined in Section 4 do not predict an interaction effect for this dependent variable. There are marginally significant negative effects of age and being in the public good treatment. The latter effect is unrelated to the task that will be executed after public sector size preferences have been measured (because this task always involves a public good, the level of which is chosen via this measurement). Instead, the dummy measures whether or not the subject participated in the public good treatment before this preference was measured, The result implies that having experienced the public good leads one to prefer a smaller public sector.

Recalling that of the three mechanisms we discussed, gross wage illusion and warm glow both predict that a larger public sector is preferred in the income tax treatment, our findings suggest that tax aversion plays a dominant role when individuals determine their preference for the size of the public sector. This supports the intuition that people prefer a larger public sector when they feel (irrationally) that someone else is paying for it. It thus offers an intuitive explanation of why right-wing politicians tend to favor levying taxes at the employees' side while left-wing politicians favor levying taxes on the employers' side – if politicians understand the preferences induced by the taxes, they could use the liability side of the tax to push preferences on the size of the public sector in the direction of their own preference.

5.2 Subjective Well-Being

Table 3 shows the coefficient estimates for subjective well-being.²⁴ We can see that participants report higher subjective well-being under an employer payroll tax. This finding is significant at the 1%-level. We thus reject the null-hypothesis $H_{sw,0}$ that there is no difference in subjective well-being under an employer payroll tax and under an income tax in favor of the alternative hypothesis (at least when the tax proceeds are lost, i.e. when the tax money is not invested in a public good).

Table 3: Results for subjective well-being (sw)

	Coefficient	Std. error
Intercept	23.81***	(3.80)
Age	-0.04	(0.16)
Gender (1=male)	1.54**	(0.78)
Studies (1=econ+science)	-0.61	(0.87)
lab experience (1=yes)	-1.79	(1.17)
Tax (1=income tax)	-2.80***	(1.08)
PG (1=public good)	-0.62	(1.05)
Tax * PG	3.09**	(1.51)

Notes: The dependent variable in this regression is subjective well-being (variable sw). Individual outcomes are integers between 4 and 36, larger numbers represent higher subjective well-being. 1 observation is left censored, 197 are uncensored, none are right censored. *, **, and *** stand for significance at the 10-, 5-, and 1-percent level.

The interaction effect between the tax treatment dummy and the public good dummy is positive and significantly different from zero (at the 5%-level). This means that the effect of larger subjective well-being under an employer payroll tax is weakened or even reversed when tax proceeds are used beneficially (for our parameters, the two effects are of more or less equal magnitude).

Looking back at our mechanisms, the results suggest that with nothing in return, the effect of tax aversion outweighs the effect of gross wage illusion on subjective well-being (because the coefficient of the tax dummy is significantly negative). Warm glow only plays a role when the tax proceeds are contributed to a public good. It predicts higher subjective-well being under an income tax. The findings for the interaction effect thus support a warm-glow effect. Together, these results on subjective well-being can be explained very intuitively by the mechanisms driving them. Tax payers do not like

²⁴Recall that the measurement of subjective well-being is the sum of four responses to the manikin scale we used. Consideration of the four separate self-reports shows very consistent behavior and no indication of trends, however.

the feeling of money being taken away, which makes them less happy under an income tax. However, the more productively tax revenue is used for a collective good, the less they mind the tax being deducted.

5.3 Labor Supply and Job Performance

The measures on labor supply and job performance we use can be divided into two parts. The first part consists of two measures of labor supply at the extensive margin (once extracted before the randomization and experiencing the work task and once after). These measures capture the willingness to bear a cost in order to work. We have argued that this is indeed a relevant measure of labor supply at the extensive margin. The extensive labor supply decisions are taken outside of the regular work environment in our experiment, but arguably also in the world outside of the laboratory.²⁵

The second part of our measures concern labor supply at the intensive margin and job performance. These correspond to decisions/performance made while working. In our experiment, subjects are solving problems while time is running down and while we are measuring how well they do the task. If they get annoyed, stressed, or bored, or if they no longer deem the task profitable, they can stop supplying labor by quitting the task for the remainder of the current round. Labor supply at the intensive margin and job performance are of course related; participants spending more time solving problems are likely to solve more of them. They also differ substantially different from the two measures of labor supply at the extensive margin. For this reason, we will separate the discussion and first discuss labor supply at the extensive margin and then labor supply at the intensive margin and job performance.²⁶

5.3.1 Labor Supply at the Extensive Margin

Table 4 shows the results for labor supply at the extensive margin. We present separate results for the two measures. The first, shows that, *ceteris paribus*, an income tax leads

²⁵Outside of the laboratory, this decision will often be made when people know which kinds of job they are applying to, but do not really know what they can precisely expect. This corresponds to our first measure – participants have read the description of the work task, know the payments, and have correctly answered the test questions to make sure that they understand the task, but they have not yet experienced the work task. Our second measure is similar, but measured after four rounds of work (in the outside world, this corresponds to the decision to apply to a similar job, if available, after termination of a job). Again, the decision is the willingness to bear a cost for participation.

²⁶The two measures at the extensive margin are positively correlated (0.380), as are performance and supply at the intensive margin (0.804). The four ‘cross-correlations’ are lower, varying between 0.125, and 0.244.

to higher willingness to participate in the labor market than an employer payroll tax. The coefficient of the tax dummy is significant at the 5%-level. We therefore reject the null hypothesis $H_{le1,0}$ stating that the tax liability side makes no difference for this measure of labor supply. The coefficient of the interaction between tax dummy and public good dummy is not significantly different from zero.

Table 4: Results for labor supply at the extensive margin

	Measure 1 (<i>le1</i>)		Measure 2 (<i>le2</i>)	
	Coefficient	Std. error	Coefficient	Std. error
Intercept	1378.04***	(220.72)	1282.63***	(281.23)
Age	0.02	(9.37)	-1.09	(11.95)
Gender (1=male)	142.80***	(45.28)	96.51*	(57.70)
Studies (1=econ+science)	10.90	(49.92)	-77.71	(64.51)
lab experience (1=yes)	-84.08	(67.99)	-48.44	(85.97)
Tax (1=income tax)	125.77**	(62.39)	94.51	(79.18)
PG (1=public good)	12.42	(61.95)	-35.93	(78.16)
Tax * PG	-85.91	(87.84)	-3.24	(110.95)

Notes: The dependent variables in these regressions are the two measures of labor supply at the extensive margin (preplay variable *le1* and postplay variable *le2*). Individual outcomes are integers between 1056 and 2400, larger numbers represent higher labor supply. For variable *le1*, 20 observations are left censored, 216 are uncensored, and 4 are right censored (for the first variable there are more observations, because the preplay measure was elicited before the randomization). For variable *le2*, 42 observations are left censored, 152 are uncensored, and 6 are right censored. *, **, and *** stand for significance at the 10-, 5-, and 1-percent level.

For the second measure, the general pattern is the same. The coefficient of the income tax dummy is positive (and in magnitude larger than one standard deviation of the estimate), but not significantly different from zero. The coefficient of the interaction term between tax and public good condition is close to zero. We therefore cannot reject hypothesis $H_{le2,0}$. Nevertheless, after rejecting $H_{le1,0}$ and seeing the same pattern for the second measure, noting that we have lower power to detect differences for the second measure (which is only available for employees, because it is extracted after the randomization), we interpret the overall results on labor supply at the extensive margin as evidence that this labor supply is higher under an income tax than under an employer payroll tax.

It seems worth noting that for both measures, men are willing to pay significantly more to participate in an extra round than women. This may be due to (over-)confidence, i.e., men think that they can do the job better than women think that they can do job (men indeed perform slightly better, although this difference is not statistically significant,

we will look at performance in more detail below).

As for the three mechanisms, gross wage illusion seems to be dominant when participants make their decisions on labor supply at the extensive margin, because the higher gross wage in the income tax case makes subjects more willing to work. Tax aversion and warm glow seem to play at most a minor role (tax aversion weakens the effect of gross wage illusion; warm glow would lead to a positive interaction term between income tax and public good).

5.3.2 Labor Supply at the Intensive Margin and Job Performance

Table 5 shows the results for labor supply at the intensive margin and job performance. In both cases, the coefficient of the income tax dummy is negative, but relatively small and insignificant. Not surprisingly, subjects with a relatively high math content in their studies (economics and science) are better at solving the problems than other students – the coefficient of the dummy indicating the field of study is positive and significantly different from zero at the 5%-level. For both outcome variables, the coefficient of the interaction of tax and public good condition is positive and much larger than the negative coefficient of the tax type dummy. For labor supply at the intensive margin, this coefficient is marginally significant. This hints at larger labor supply at the intensive margin and job performance under an income tax when the tax proceeds are used beneficially.

Table 5: Results for labor supply at the intensive margin and job performance

	Labor supply, intensive (li)		Job performance (jp)	
	Coefficient	Std. error	Coefficient	Std. error
Intercept	1636.92***	(390.39)	18.91***	(5.88)
Age	1.61	(16.81)	-0.04	(0.25)
Gender (1=male)	-21.96	(80.57)	1.87	(1.21)
Studies (1=econ+science)	143.60	(89.72)	3.20**	(1.35)
lab experience (1=yes)	-107.49	(121.60)	-1.13	(1.81)
Tax (1=income tax)	-86.45	(111.44)	-0.53	(1.67)
PG (1=public good)	-198.75*	(108.42)	-1.87	(1.64)
Tax * PG	281.49*	(155.95)	3.14	(2.34)

Notes: The dependent variables in these regressions are labor supply at the intensive margin (li) and job performance (jp). Individual outcomes are between 0 and 1920 for li (representing time spent on the work task) and at least 0 for jp (where the number corresponds to correctly solved problems, the maximum achieved in the data is 37). For variable li , 0 observations are left censored, 158 are uncensored, and 41 are right censored. For variable jp , 4 observations are left censored and 195 are uncensored. *, **, and *** stand for significance at the 10-, 5-, and 1-percent level.

The mechanisms we discussed seem to be either balanced (gross wage illusion and tax aversion work in opposite directions) or not strong enough when subjects take these decisions ‘under stress’ (i.e. while working). The fact that the interaction term of tax and public good condition is positive and marginally significant for labor supply at the intensive margin is a further (albeit weak) indication that our mechanisms play a role in subjects’ decisions.

6 Policy Implications

The policy implications of our results are non-trivial. Our results show that who formally pays the tax affects individuals in at least three ways: their political preferences, their subjective well-being, and their labor supply at the extensive margin (with indications of similar effects at the intensive margin and performance as well; see Fochmann et al., 2010b). Under LSE, none of these should be affected. Each of the effects may have consequences for optimal policies. Simultaneously considering the three different categories would make deducing policy implications very complex and dependent on auxiliary assumptions – it is beyond the scope of this paper to develop such a large and complex optimal taxation model. Instead, we consider each of the findings in isolation and discuss their implications for optimal policies, other things equal (ordered from easiest to draw policy conclusions to least easy). Naturally, when determining actual governmental policies, many other things need to be taken into account, including collection and compliance costs, possibilities to allow for tax deductions, possibilities of introducing progression and special tax rates, and possibilities to impede evasion and avoidance. Hence, the following implications should be seen as an indication of the direction in which optimal policy may move. In spite of these caveats, we believe that it is important to consider the policy implications. Some of the effects we observe in the laboratory are strong. In the world outside of the laboratory, even very small effects are relevant, because labor taxes are extremely widespread (and e.g. deadweight losses from distortionary taxes on labor may be substantial; Feldstein, 1999).

If citizens’ subjective well-being is the government’s main concern, the policy implications of our findings are relatively straightforward. We observe that subjective well-being is higher under an employer payroll tax.²⁷ This finding thus suggests that, *ceteris paribus*, it would be better to levy taxes on the employer’s side.

²⁷In our experiment, we find large and highly significant differences in the case that tax proceeds are lost. As explained in the introduction, we view the real world as lying somewhere in between the two extremes of totally lost and extremely profitable tax proceeds. In this case, we expect subjective well-being to increase if more of the taxes are levied at the employers’ side.

Concerning labor supply, most economists would probably agree that taxes should be chosen that minimize the disruption of the price mechanism in the labor market. As labor taxes are generally thought to reduce labor supply, this implies that the tax should be chosen that maximizes labor supply (and output).²⁸ If people work more under one labor tax scheme than under the other, using this scheme brings society closer to first best and thus increases social welfare. The experimental findings concerning labor supply are that people tend to supply more labor under an income tax than under an employer payroll tax. Thus, the optimal policy for the government when considering only labor supply responses would be to rely only on income taxes.

It is less straightforward to draw policy conclusions from our results regarding preferences for the size of the public sector. To start, note that such preferences may be expressed in votes and therefore may affect the actual size of the public sector in the economy. If one assumes that there is something like an optimal size, it seems that the combination of labor taxes is optimal that induces individuals to prefer this size of the public sector. We cannot, however, draw conclusions from our finding that people prefer a larger public sector when the tax is levied on the employer's side as we cannot infer without additional assumptions what the 'objectively' optimal size of the public sector would be in the experiment. There might thus be an optimal policy choice from a normative point of view. In a positive analysis, though, it is likely that a politician will try to use the distinct taxes to induce preferences in the electorate that are closer to her own view of optimal public sector size rather than to some 'objectively' optimal size.

7 Concluding Remarks

The question whether a labor tax levied at the employees' side of the labor market and one levied at the employers' side are equivalent is relevant for policy making, political economics and optimal taxation theory. In this paper we have investigated this LSE in a controlled laboratory experiment. Specifically, we have focused on the effects of this distinct framing of otherwise equivalent taxes, arguing that framing effects provide direct evidence against tax equivalence. Our results support the claim that these duties are not equivalent. In particular, our results suggest that employees prefer a larger public sector and that subjective well-being is higher when the tax is levied on the

²⁸Assume that the government needs to raise a fixed amount of revenue to finance its expenditures and wants to do so via labor taxes. These taxes constitute a disincentive to work, such that labor supply is lower than under the first best solution, which consists of lump-sum taxes (assuming that labor supply elasticities are as usually reported (see Evers et al., 2008), i.e. that the substitution effect is larger than the income effect).

employer's side, while labor supply tends to be lower at the extensive margin.

We have also highlighted three mechanisms that may explain the framing effects we observe, to wit, gross wage illusion, tax aversion and warm glow. Our results show that all three seem to be at work, and have distinct effects on preferred public sector size, subjective well-being and labor supply.

We elaborated briefly on the policy implications of these findings, which are not trivial. The aspects we investigated lead, when considered separately, to distinct implications and it is quite possible that the optimal policy involves a mix of labor taxes levied at the employers' and employees' sides. This could explain the 'puzzle' of why these theoretically equivalent duties often exist side by side instead of governments simply adopting the duty that minimizes collection and compliance costs, even in the absence of labor market frictions. As explained above, an alternative solution to this 'puzzle' based on political-economic considerations is that politicians choose a tax mix that induces the public to prefer a size of the public sector close to the one the politicians themselves favor.

It seems clear from our results that boundedly rational behavior plays an important role in individuals' reactions to taxes. Classic optimal taxation theory, which assumes individual rationality, is thus based on an empirically shaky foundation. The development of normative optimal taxation models encompassing non-rational perception of taxes seems an important line of future research. The same holds for the incorporation of tax perception into positive models of political economics. Our results may aid in the development of such models.

References

- Anand, P. and Van Hees, M. (2006). Capabilities and achievements: An empirical study. *The Journal of Socio-Economics*, 35(2):268–284.
- Andreoni, J. (1990). Impure altruism and donations to public goods: a theory of warm-glow giving. *The economic journal*, 100(401):464–477.
- Atkinson, A. and Stiglitz, J. (1980). *Lectures on public economics*. Economics handbook series. McGraw-Hill Book Co.
- Bassi, M. (2010). Mirrlees meets laibson: Optimal income taxation with bounded rationality. *CSEF Working Papers*.

- Becker, G. M., Degroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Systems Research and Behavioral Science*, 9:226–232.
- 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.
- Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59.
- Charness, G. and Kuhn, P. (2011). Lab labor: What can labor economists learn from the lab? *Handbook of labor economics*, 4:229–330.
- Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77.
- Conlisk, J. (1996). Why bounded rationality? *Journal of economic literature*, pages 669–700.
- Cox, J. C., Rider, M., and Sen, A. (2012). Tax incidence: Do institutions matter? an experimental study. Technical report, Experimental Economics Center, Andrew Young School of Policy Studies, Georgia State University.
- Crosen, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74.
- De Bartolome, C. A. M. (1995). Which tax rate do people use: Average or marginal? *Journal of Public Economics*, 56(1):79–96.
- Evers, M., De Mooij, R., and Van Vuuren, D. (2008). The wage elasticity of labour supply: A synthesis of empirical estimates. *De Economist*, 156:25–43.
- Falk, A. and Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences.
- Fehr, E. and Goette, L. (2007). Do workers work more if wages are high? evidence from a randomized field experiment. *American Economic Review*, 97(1):298–317.
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868.
- Feldstein, M. (1999). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics*, 81(4):674–680.

- Ferrer-i Carbonell, A. (2013). Happiness economics. *SERIEs*, pages 1–26.
- Finkelstein, A. (2009). E-ztax: Tax salience and tax rates. *The Quarterly Journal of Economics*, 124(3):969–1010.
- Fochmann, M., Kiesewetter, D., Blaufus, K., Hundsdorfer, J., and Weimann, J. (2010a). Tax Perception - An Empirical Survey. *Arqus Discussion Paper No. 99*.
- Fochmann, M., Kiesewetter, D., and Sadrieh, A. (2012). Investment behavior and the biased perception of limited loss deduction in income taxation. *Journal of Economic Behavior and Organization*, 81(1):230 – 242.
- Fochmann, M. and Weimann, J. (2011). The effects of tax salience and tax experience on individual work efforts in a framed field experiment.
- Fochmann, M., Weimann, J., Blaufus, K., Hundsdorfer, J., and Kiesewetter, D. (2010b). Grosswage illusion in a real effort experiment. *FEMM Working Papers*.
- Hayashi, A., 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*.
- Holm, P., Honkapohja, S., and Koskela, E. (1994). A monopoly-union model of wage determination with capital and taxes: An empirical application to the finnish manufacturing. *European Economic Review*, 38(2):285–303.
- Hossain, T. and List, J. A. (2012). The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science*, 58(12):2151–2167.
- Irtel, H. (2007). Pxlabs: The psychological experiments laboratory [online]. version 2.1.11. mannheim (germany): University of mannheim. [cited 25 april 2011]. available from <<http://www.pxlabs.de>>.
- Kahneman, D. and Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *The journal of economic perspectives*, 20(1):3–24.
- Kallbekken, S., Kroll, S., and Cherry, T. L. (2010). Pigouvian tax aversion and inequity aversion in the lab. *Economics Bulletin*, 30(3):1914–1921.
- Koskela, E. and Schöb, R. (1999). Does the composition of wage and payroll taxes matter under nash bargaining? *Economics Letters*, 64(3):343–349.

- Lang, P. J. (1985). *The Cognitive Psychophysiology of Emotion: Anxiety and Anxiety Disorders*. Lawrence Erlbaum, Hillsdale, NJ.
- Lehmann, E., Marical, F., and Rioux, L. (2013). Labor income responds differently to income-tax and payroll-tax reforms. *Journal of Public Economics*, 99:66–84.
- McCaffery, E. and Slemrod, J. (2006). *Behavioral Public Finance*. G - Reference, Information and Interdisciplinary Subjects Series. Russell Sage.
- McCaffery, E. J. and Baron, J. (2006). Thinking About Tax. *Psychology, Public Policy, and Law*, 12(1):106–135.
- Mullainathan, S., Schwartzstein, J., and Congdon, W. J. (2012). A reduced-form approach to behavioral public finance. *Annu. Rev. Econ.*, 4(1):511–540.
- Muysken, J., Van Veen, T., and ERIK, D. R. (1999). Does a shift in the tax burden create employment? *Applied Economics*, 31(10):1195–1205.
- Oswald, A. J. (1985). The economic theory of trade unions: An introductory survey. *The Scandinavian Journal of Economics*, pages 160–193.
- Picard, P. and Toulemonde, E. (2001). On the equivalence of taxes paid by employers and employees. *Scottish Journal of Political Economy*, 48(4):461–470.
- Pissarides, C. A. (2000). *Equilibrium unemployment theory*. MIT press.
- Rasmussen, B. (1997). Non-equivalence of employment and payroll taxes in imperfectly competitive labor markets. *Univ. of Aarhus Dept. of Economics WP*, 22.
- Riedl, A. and Tyran, J.-R. (2005). Tax liability side equivalence in gift-exchange labor markets. *Journal of Public Economics*, 89(11-12):2369–2382.
- Robins, P. K. (1985). A comparison of the labor supply findings from the four negative income tax experiments. *Journal of Human Resources*, pages 567–582.
- Saez, E., Matsaganis, M., and Tsakloglou, P. (2012). Earnings determination and taxes: Evidence from a cohort-based payroll tax reform in greece. *The Quarterly Journal of Economics*, 127(1):493–533.
- Sausgruber, R. and Tyran, J.-R. (2005). Testing the mill hypothesis of fiscal illusion. *Public Choice*, 122:39–68. 10.1007/s11127-005-3992-4.

- Sausgruber, R. and Tyran, J.-R. (2011). Are we taxing ourselves?: How deliberation and experience shape voting on taxes. *Journal of Public Economics*, 95(1-2):164 – 176.
- Schokkaert, E. (2007). Capabilities and satisfaction with life. *Journal of Human Development*, 8(3):415–430.
- Schram, A. (2005). Artificiality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology*, 12(2):225–237.
- Shapiro, C. and Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *The American Economic Review*, pages 433–444.
- Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 69(1):99–118.
- Stiglitz, J. (2000). *Economics of the Public Sector*. Norton.
- Ullmann, R. and Watrin, C. (2008). Comparing direct and indirect taxation: The influence of framing on tax compliance. *European Journal of Comparative Economics*, 5(1):23–56.
- Weber, M. (2011). Income tax or employer’s contribution - an experiment on the perception of ‘economically equivalent’ duties. Mphil-thesis, Tinbergen Institute.

A Appendix (for Online Publication): Instructions, Test Questions, and Questionnaire

We reproduce here the on-screen instructions and test questions. They are in the same order as they appear during the experiment (in between are the tasks as described in Section 3). We have split the instructions for the different treatments where relevant. Aside from the on-screen instructions, subjects received a one-page summary which has no new information. Text that is bold and/or italic on the screen is also so in this appendix. Multiple choice test questions are shown here with empty squares for the possible answers, test questions where an input is required are shown with an empty circle below the question. Note that it is only possible to proceed from a page of test questions after all questions on this page have been answered correctly. If not, the message “You did not answer all questions correctly. Take another look at the instructions or raise your hand if you need help.” is displayed, without telling the participant which question(s) has/have been answered incorrectly. It is thus necessary for the participants to fully understand the instructions and to know the answers to the questions rather than just clicking through them, also for multiple choice questions. In Section A.2 we reproduce the questionnaire for the subjects in the role of employees.

A.1 Instructions and Test Questions

FIRST INSTRUCTION SCREEN, ALL TREATMENTS

Welcome to this experiment!

Depending on your decisions and the decisions of other participants in today’s experiment, you can earn money. You will be paid privately at the end of the experiment. Your earnings will not be revealed to anyone. This is an anonymous experiment; your decisions will only be linked to your station id and not to your name in any way. The experiment will take approximately 2 hours.

This experiment may involve gains and losses. Whether the possibility of a loss occurs is completely determined by your own decisions. It is thus possible (though unlikely) that you make a negative amount of money in the experiment. In this case, your losses will be deducted from your earnings and from the show-up fee.

This experiment is composed of different parts. You will receive instructions for the different parts before they begin. After the instructions you will have to answer a few test questions before you can continue.

Please read all instructions carefully. You are not allowed to speak with other participants or to communicate with them in any other way.

Payments in most parts are in points, but there are also payments in Euro. At the end of the experiment points will be exchanged into Euro at the exchange rate of 600 points = 1 Euro.

By showing up, you have already earned 7 Euro. This show-up fee will be added to your earnings from the experiment.

If you want to ask a question, please raise your hand and someone will come to your desk.

SECOND INSTRUCTION SCREEN, ALL TREATMENTS

In this part you will be randomly divided into groups of six people each. One person in each group will be randomly determined to be "employer" the other five persons will be "employees". The group composition will not change during the whole experiment.

This part consists of four rounds. Each will last for 8 minutes. In each round, the employer hires the five employees to perform a work task.

What the employees have to do:

You (employees only) will see two matrices on the screen. Each matrix has 10 rows and 10 columns and is filled with randomly generated numbers. Your job is to find the largest number in each of the matrices and then to add them up. You are allowed to use the pocket calculators on your table. For each correct solution, the employer will pay you a wage. This wage becomes lower for each new problem you face.

After entering your answer you will be told whether your answer is correct or not (please note that the time will continue to run while you see this result). Subsequently, irrespective of whether your answer is correct or incorrect, a new pair of matrices will appear. This means that for each pair, you have only one attempt to provide the correct answer.

Instead of trying to solve problems you can also choose to use a "fixed payment option". You will see a button saying "Go to fixed payment option" at the bottom right of your screen. If you click on this button you will receive a fixed payment for each second remaining in the round; while the time ticks away you will see a basically empty screen. You will then not be able to solve any more problems in this round (thus you cannot go back and forth between problem solving and the fixed payment option).

At the top of the screen you can see how many of your answers in this round were right and wrong. Here, you can also see your total wage for this round. You will see the time that remains in this round in the upper right corner of the screen. You will also always see the wage you will receive if your next problem is solved correctly. You can try to solve at most thirty of these problems per round (which will be much more than anyone can actually solve).

What the employers have to do: While the employees are working, the employers do not need to do anything.

After each round there will be a short one-click questionnaire. Please answer this question as honestly and accurately as you can.

FIRST PART OF THE THIRD INSTRUCTION SCREEN

TREATMENT *EN*

Payments:

For each correct addition an employee makes, his/her employer (the employer in the same group) will receive a certain number of points. From these points, the employer pays a wage to the employee. Aside from the wage paid to the employee, the employer will also pay (to the experimenter) an amount equal to 66.7% of the employee's wage. The employer thus has to pay more than only the employees' wages.

The wage an employee receives for a correct addition depends on how many problems this employee has already attempted to solve in this round. If the first attempt is correct, the employee's wage is 280.8 points. For each subsequent attempt the wage decreases by 23.4 points, but it is never smaller than zero. As employee you will always see on the screen the wage you will receive if the next addition is correct. For an incorrect addition neither the employer nor the employee will receive any points. The amount the employer receives for a correct addition of an employee is such that he/she will in the end always keep as a profit 49.8 points per correct addition.

Instead of solving problems, an employee can also go to the fixed payment option. If an employee goes to the fixed payment option, he/she receives 2.2 points per second that is still remaining in this round. The employer earns nothing from the fixed payment option, he/she only earns points for correctly solved problems.

TREATMENT *IN*

Payments:

For each correct addition an employee makes, his/her employer (the employer in the same group) will receive a certain number of points. From these points, the employer pays a wage to the employee. From the employee's wage, 40% will be deducted, such that an employee in the end only receives 60% of his/her wage.

The wage an employee receives for a correct addition depends on how many problems this employee has already attempted to solve in this round. If the first attempt is correct, the employee's wage is 468 points. For each subsequent attempt the wage decreases by 39 points, but it is never smaller than zero. As employee you will always see on the screen the wage you will receive if the next addition is correct. For an incorrect addition neither the employer nor the employee will receive any points. The amount the employer receives for a correct addition of an employee is such that he/she will in the end always keep as a profit 49.8 points per correct addition.

Instead of solving problems, an employee can also go to the fixed payment option. If an employee goes to the fixed payment option, he/she receives 2.2 points per second that is still remaining in this round. From these points nothing is deducted. The employer earns nothing from the fixed payment option, he/she only earns points for correctly solved problems.

TREATMENT *EP*

Payments:

For each correct addition an employee makes, his/her employer (the employer in the same group) will receive a certain number of points. From these points, the employer pays a wage to the employee. Aside from the wage paid to the employee, the employer will also pay an amount that equals 66.7% of the employee's wage. The proceeds will be put into a common fund. Each group has its own common fund. At the end of the experiment, the total number of points in the common fund will be multiplied by 1.3 and then distributed equally among all employees in this group.

The wage an employee receives for a correct addition depends on how many problems this employee has already attempted to solve in this round. If the first attempt is correct, the employee's wage is 239.32 points. For each subsequent attempt the wage decreases by 19.94 points, but it is never smaller than zero. As employee you will always see on the screen the wage you will receive if the next addition is correct. For an incorrect addition neither the employer nor the employee will receive any points. The amount the employer receives for a correct addition of

an employee is such that he/she will in the end always keep as a profit 49.8 points per correct addition.

Instead of solving problems, an employee can also go to the fixed payment option. If an employee goes to the fixed payment option, he/she receives 2.2 points per second that is still remaining in this round. The employer earns nothing from the fixed payment option, he/she only earns points for correctly solved problems (thus, the employer will not contribute points to the common fund for the time an employee uses the fixed payment option).

TREATMENT *IP*

Payments:

For each correct addition an employee makes, his/her employer (the employer in the same group) will receive a certain number of points. From these points, the employer pays a wage to the employee. From the employee's wage, 40% will be deducted and put into a common fund. Each group has its own common fund. At the end of the experiment, the total number of points in the common fund will be multiplied by 1.3 and then distributed equally among all employees in this group.

The wage an employee receives for a correct addition depends on how many problems this employee has already attempted to solve in this round. If the first attempt is correct, the employee's wage is 398.87 points. For each subsequent attempt the wage decreases by 33.24 points, but it is never smaller than zero. As employee you will always see on the screen the wage you will receive if the next addition is correct. For an incorrect addition neither the employer nor the employee will receive any points. The amount the employer receives for a correct addition of an employee is such that he/she will in the end always keep as a profit 49.8 points per correct addition.

Instead of solving problems, an employee can also go to the fixed payment option. If an employee goes to the fixed payment option, he/she receives 2.2 points per second that is still remaining in this round. From these points nothing is deducted. The employer earns nothing from the fixed payment option, he/she only earns points for correctly solved problems.

SECOND PART OF THE THIRD INSTRUCTION SCREEN, ALL TREATMENTS

Before the rounds start:

Before the problem solving rounds start you will be told whether you are employer or employee. Before this, you will be asked how much you would be willing to pay in order to work (and get paid) for an extra round in case that you are employee. There will be a random mechanism deciding whether or not this extra round will actually take place. If the round does not take place or if you are employer, the number you state will not have any consequences for you. This extra round will take place (if it does take place) at the end of the experiment. It will be the same as the regular rounds, just the number of people that work in your group may be different, simply because some may not be working in this extra round. Also the payments will be the same as in the regular rounds.

If the random process we use determines that the extra round will in fact take place, the following occurs. The price of participating in this extra round is randomly determined and lies between 1056 and 2400 points. If this price is lower than the number you state, the price will be deducted from your earnings and you will work and get paid for an extra round. If the price is higher than the number you state, nothing will be deducted from your earnings and you will not work for an extra round. Note that we will not pay anyone until the whole experiment is

finished. Note that you will have to state an amount of at least 1056 points (this is the amount of points you get for a round if you immediately switch to the fixed payment option).

Thus, you will not pay for the extra round the amount you state that you are willing to pay. You (if you are employee) either pay the randomly determined price – in case that this price is lower than the number you state – or you pay nothing and do not play an extra round.

If you are employer you will again not have to do anything in the extra round, but you will earn the same amount of points per correct addition of your employees as in the regular rounds.

FIRST TEST QUESTION SCREEN, ALL TREATMENTS

Before the experiment starts, we will ask you some questions to check your understanding. You can return to the instructions by clicking on the menu at the top of the screen.

When do the matrices that you see on the screen change?

- After you have entered the correct solution.
- After you have entered a number, irrespective of whether it is correct or not.
- After you have entered the correct solution or after 1 minute.

How often can you go back from the fixed payment option to solving problems in one round?

- Never.
- Once.
- As often as you like.

How many minutes does each of the rounds last?

○

The matrices that you will see during the experiment will be much larger. For now, assume that one of the matrices on your screen consists of the numbers 19, 23, 41, 16, 25, 30, 12, 29, 22 and the other matrix consists of the numbers 31, 36, 20, 15, 28, 38, 17, 19, 31. What would be the correct number to enter?

○

How large will each of the matrices that you see during the experiment be?

- 5 rows and 5 columns (25 numbers)
- 10 rows and 10 columns (100 numbers)
- 20 rows and 20 columns (400 numbers)

FIRST PART OF THE SECOND TEST QUESTION SCREEN

TREATMENT *EN*

Imagine that you are employee. On top of your wage for each correct addition, how much extra will be taken away from your employer (expressed as a percentage of your wage)?

- 33.3
- 66.7
- 50

TREATMENT *IN*

Imagine that you are employee. How much of your wage for each correct addition will be taken away from you (as a percentage of your wage)?

- 22
- 40
- 50

TREATMENT *EP*

Imagine that you are employee. On top of your wage for each correct addition, how much extra will be taken away from your employer and contributed to the common fund (expressed as a percentage of your wage)?

- 33.3
- 66.7
- 50

TREATMENT *IP*

Imagine that you are employee. How much of your wage for each correct addition will be taken away from you and contributed to the common fund (as a percentage of your wage)?

- 22
- 40
- 50

SECOND PART OF THE SECOND TEST QUESTION SCREEN, ALL TREATMENTS

If you are employer, how many points will you earn for each second that one of your employees chooses the fixed payment option?

-

If you are employee, how many points will you earn per *minute* that you choose the fixed payment option?

-

THIRD TEST QUESTION SCREEN, ALL TREATMENTS

As you know, before you will start the experiment you will be asked how much you would be willing to pay to participate in an extra round. The test questions now concern this statement. Please note that any numbers here are only meant to serve as an example. The content is not informative on what you should decide in the experiment.

Imagine that you state an amount of 1300 points and that the randomly drawn price is 1344 points. What will happen?

- 1300 points are deducted from your earnings, but you will not participate in the extra round.
- 1344 points are deducted from your earnings and you will participate in the extra round, if it takes place.
- 1300 points are deducted from your earnings and you will participate in the extra round, if it takes place.
- No points are deducted from your earnings and you will not participate in the extra round, if it takes place.

Imagine that the randomly drawn price is 1257.3 points and that it is lower than the amount you state. Imagine that the extra round takes place and the you go immediately to the fixed payment option. Will you then earn more in the extra round than the price you pay?

- Yes.
- No.

Imagine that the amount you state is 1100 points and the randomly determined price is 1429 points. Will you be able to leave the experiment earlier in this case than if you had stated 2000 points?

- Yes.
- No.
- That depends on the choices of the other participants in my group.

SUMMARY OF THE INSTRUCTIONS FOR THE FIRST PART, ALL TREATMENTS

Summary of instructions

Please note that you have a short summary of instructions including the payments lying on your desk.

You will be randomized into employees and employers.

If you are employee you solve problems that consist of finding the largest number in each of two matrices and adding the numbers up. You will receive a wage for the correct additions. Instead of working you can also choose a fixed payment option. For the payments see the sheet on your desk.

If you are employer you hardly have to do anything.

Before beginning you will be asked how much you would at most be willing to pay to participate in an extra round at the end of the experiment if you are employee. You will not pay the amount you enter, but either a randomly drawn price if this price is lower than the amount you enter or nothing.

If you want to ask a question, please raise your hand and someone will come to your desk.

INSTRUCTIONS AFTER THE REGULAR WORK ROUNDS

TREATMENTS *EN* AND *IN*

Welcome to the next part of the experiment

There is again a chance to participate in an extra round at the end of the experiment. The round will again be the same as the regular work rounds. Also the payments will be the same .

You will be asked to state the maximum price you are willing to pay to participate in this extra round. If the randomly drawn price (between 1056 and 2400) is lower than what you stated you pay this price and participate in the extra round. (The rules are thus the same as in the beginning of the experiment.)

Please note that you will not be able to finish the experiment early if you do not participate in the extra round. Please also note that it is uncertain whether or not the extra round will take place. In fact, either the extra round based on your maximum price decision before the regular rounds will take place or an extra round based on your next maximum price decision will take place. Therefore, if an extra round will take place based on your statement from before the regular rounds, this extra round based on the statement now will not take place and if no extra round will take place based on your statement from before the regular rounds, this extra round based on the statement now will take place.

This means that from your two maximum prices, only one will be considered.

If you have any questions, please raise your hand and one of us will come to your desk to answer the question.

TREATMENTS *EP* AND *IP*

Welcome to the next part of the experiment

There is again a chance to participate in an extra round at the end of the experiment. The round will again be the same as the regular work rounds. Also the payments will be the same (the common fund will be distributed among all employees of the group, also among the ones that do not participate).

You will be asked to state the maximum price you are willing to pay to participate in this extra round. If the randomly drawn price (between 1056 and 2400) is lower than what you stated you pay this price and participate in the extra round. (The rules are thus the same as in the beginning of the experiment.)

Please note that you will not be able to finish the experiment early if you do not participate in the extra round. Please also note that it is uncertain whether or not the extra round will take place. In fact, either the extra round based on your maximum price decision before the regular rounds will take place or an extra round based on your next maximum price decision will take place. Therefore, if an extra round will take place based on your statement from before the regular rounds, this extra round based on the statement now will not take place and if no extra round will take place based on your statement from before the regular rounds, this extra round based on the statement now will take place.

This means that from your two maximum prices, only one will be considered.

If you have any questions, please raise your hand and one of us will come to your desk to answer the question.

INSTRUCTIONS BEFORE THE PREFERENCE FOR PUBLIC SECTOR SIZE ELICITATION

TREATMENT *EN*

Welcome to the next part of the experiment

There will be one final extra round in the end of the experiment. Everyone will participate in this final round. This round will be a bit different from the regular rounds.

In this round, the amount that the employer pays beside the wage is no longer paid to the experimenter. Instead, it is put into a common fund. Each group has its own common fund. At the end of the experiment, the total number of points in the common fund will be multiplied by a certain factor and then distributed equally among all employees in this group.

Some of the parameters will now be decided on by the employees. You will be asked to state which set of parameters you would prefer for this extra round of the experiment. Later, the preferred set of parameters of one employee of your group will be chosen randomly and the additional extra round will take place with this set of parameters.

You will see a slider (a button that you can move horizontally) to determine your preferred set of parameters. If you move the slider from left to right, the factor with which the points in the common fund will be multiplied decreases. At the same time, the number of points that your employer contributes to the common fund increases. You are asked to choose the position of the slider that you prefer. Please note that the profit of the employer is held constant. Therefore, higher contributions to the common fund will lead to lower wages.

You can look at this problem in the following way (although this is not the only possible way). You would like the multiplication factor to be big, because it will be multiplied with the points in the common fund before these points are distributed among the employees of your group. But if you choose the slider to be as far to the left as possible, nothing will be contributed to the common fund at all and no one can profit from the large multiplication factor. As explained, you should choose the combination of the multiplication factor and the size of the contributions to the common fund that you prefer most.

If you have any questions, please raise your hand and one of us will come to your desk.

TREATMENT *IN*

Welcome to the next part of the experiment

There will be one final extra round in the end of the experiment. Everyone will participate in this final round. This round will be a bit different from the regular rounds.

Now the points that are deducted from your wage are no longer "lost". Instead, they are put into a common fund. Each group has its own common fund. At the end of the experiment, the total number of points in the common fund will be multiplied by a certain factor and then distributed equally among all employees in this group.

Some of the parameters will now be decided on by the employees. You will be asked to state which set of parameters you would prefer for this extra round of the experiment. Later, the preferred set of parameters of one employee of your group will be chosen randomly and the additional extra round will take place with this set of parameters.

You will see a slider (a button that you can move horizontally) to determine your preferred set of parameters. If you move the slider from left to right, the factor with which the points in the common fund will be multiplied decreases. At the same time, the number of points that is deducted from your wage and contributed to the common fund increases. You are asked to choose the position of the slider that you prefer.

You can look at this problem in the following way (although this is not the only possible way). You would like the multiplication factor to be big, because it will be multiplied with the points in the common fund before these points are distributed among the employees of your group. But if you choose the slider to be as far to the left as possible, nothing will be contributed to the common fund at all and no one can profit from the large multiplication factor. As explained, you should choose the combination of the multiplication factor and the size of the contributions to the common fund that you prefer most.

If you have any questions, please raise your hand and one of us will come to your desk.

TREATMENT *EP*

Welcome to the next part of the experiment

There will be one final extra round in the end of the experiment. Everyone will participate in this final round. This round will be a bit different from the regular rounds.

Some of the parameters will now be decided on by the employees. You will be asked to state which set of parameters you would prefer for this extra round of the experiment. Later, the preferred set of parameters of one employee of your group will be chosen randomly and the additional extra round will take place with this set of parameters.

You will see a slider (a button that you can move horizontally) to determine your preferred set of parameters. If you move the slider from left to right, the factor with which the points in the common fund will be multiplied decreases (of course, this refers only to the points contributed to the common fund in this extra round). At the same time, the number of points that your employer contributes to the common fund increases. You are asked to choose the position of the slider that you prefer. Please note that the profit of the employer is held constant. Therefore, higher contributions to the common fund will lead to lower wages.

You can look at this problem in the following way (although this is not the only possible way). You would like the multiplication factor to be big, because it will be multiplied with the points in the common fund before these points are distributed among the employees of your group. But if you choose the slider to be as far to the left as possible, nothing will be contributed to the common fund at all and no one can profit from the large multiplication factor. As explained, you should choose the combination of the multiplication factor and the size of the contributions to the common fund that you prefer most.

If you have any questions, please raise your hand and one of us will come to your desk.

TREATMENT *IP*

Welcome to the next part of the experiment

There will be one final extra round in the end of the experiment. Everyone will participate in this final round. This round will be a bit different from the regular rounds.

Some of the parameters will now be decided on by the employees. You will be asked to state which set of parameters you would prefer for this extra round of the experiment. Later, the preferred set of parameters of one employee of your group will be chosen randomly and the additional extra round will take place with this set of parameters.

You will see a slider (a button that you can move horizontally) to determine your preferred set of parameters. If you move the slider from left to right, the factor with which the points in the common fund will be multiplied decreases (of course, this refers only to the points contributed to the common fund in this extra round). At the same time, the number of points that is deducted from your wage and contributed to the common fund increases. You are asked to choose the position of the slider that you prefer.

You can look at this problem in the following way (although this is not the only possible way). You would like the multiplication factor to be big, because it will be multiplied with the points in the common fund before these points are distributed among the employees of your group. But if you choose the slider to be as far to the left as possible, nothing will be contributed to the common fund at all and no one can profit from the large multiplication factor. As explained, you should choose the combination of the multiplication factor and the size of the contributions to the common fund that you prefer most.

If you have any questions, please raise your hand and one of us will come to your desk.

FIRST PART OF THE TEST QUESTION SCREEN BEFORE THE PREFERENCE FOR PUBLIC SECTOR SIZE ELICITATION, ALL TREATMENTS

If you go to the fixed payment option instead of solving problems, will this increase the number of points in the common fund in any way?

- Yes.
- No.
- That depends on the set of parameters chosen.

If your preferred position of the slider is all the way to the left and your choice is the one that is randomly picked, how many points will you receive for each correct addition by one of the other employees in your group?

○

SECOND PART OF THE TEST QUESTION SCREEN BEFORE THE PREFERENCE FOR PUBLIC SECTOR SIZE ELICITATION

TREATMENTS *EN* AND *EP*

Does the gross number of points that your employer receives for each correct addition (i.e. the number before he/she pays the wages and the contributions to the common fund) depend on the set of parameters chosen?

- Yes.
- No.

TREATMENTS *IN* AND *IP*

Does the gross number of points that your employer receives for each correct addition (i.e. the number before he/she pays the wages) depend on the set of parameters chosen?

- Yes.

No.

THIRD PART OF THE TEST QUESTION SCREEN BEFORE THE PREFERENCE FOR PUBLIC SECTOR SIZE ELICITATION, ALL TREATMENTS

Does the profit your employer keeps in the end from each correct addition by one of his/her employees depend on the set of parameters chosen?

Yes.

No.

INTRODUCTION OF THE LOTTERIES CONCERNING THE LOSS AVERSION ELICITATION, ALL TREATMENTS

Welcome to the next part of the experiment

In this part you will be presented lotteries A and B.

*A: Play the following lottery **once**.*

Win 4 Euro with probability one half, lose 2.5 Euro with probability one half.

*B: Play the following lottery **six times in a row**.*

Win 4 Euro with probability one half, lose 2.5 Euro with probability one half.

You can state for each of the two lotteries whether or not you would like to participate in them. Only one of the two lotteries will actually be played out, and it will only be played out for you if you decided to participate in it. It will later be randomly decided whether lottery A or lottery B will be played out.

If you decided to participate in the lottery that is chosen, it will be played out and you will receive the earnings. If you decided not to participate in this lottery you will neither earn nor lose any money.

If you have any questions, please raise your hand and one of us will come to your desk.

TEST QUESTIONS BEFORE CHOOSING THE LOTTERIES, ALL TREATMENTS

Please answer the following questions. (Please note that you should state decimals with a point, not with a comma.)

If you choose not to participate in either of the lotteries, how much money will you earn?

○ Euro

If you state that you would like to participate in both lotteries and lottery A is chosen randomly to be played out, how much money can you lose in the worst case scenario?

○ Euro

If you state that you would like to participate in lottery B and lottery B is chosen randomly to be played out how much money can you win in the best case scenario?

○ Euro

A.2 Questionnaire

ALL TREATMENTS

Questionnaire

The experiment has almost ended. Please fill out the following questionnaire. When you have finished, you will return to the summary of your earnings. When everyone has finished with the questionnaire we will start paying you and you may leave.

Gender:

- Male
- Female

Age:

o

Have you participated in a CREED experiment before?

- No
- Yes, once
- Yes, more than once

Department where you study:

- UVA – Faculty of Economics and Business
- UVA – Faculty of Social and Behavioural Sciences - Psychology
- UVA – Faculty of Social and Behavioural Sciences - non psychology
- UVA – Faculty of Science
- UVA – IIS: beta gamma bachelor
- UVA – Faculty of Law
- UVA - Faculty of Humanities
- UVA - Faculty of Medicine
- UVA - Faculty of Dentistry
- Another university
- A Dutch 'hogeschool' (HBO)
- Different

Did you sometimes switch to the fixed payment option? If yes, how did you decide when to switch?

○

Were there things you did not understand completely/correctly during the experiment? If yes, please state which parts. You can also leave any other comment here, if there is something you think we might be interested in knowing.

○

Consider the socio-economic system in the Netherlands. What do you think about the public sector? It is...

- Much too big
- Too big
- About right
- Too small
- Much too small
- Don't know

B Appendix (for Online Publication): Screenshots

Figure 6 shows a screenshot of an employee's computer during the work task (taken in treatment *IN*). Figure 7 shows the subjective well-being self-assessment using the SAM-V-9 self-assessment manikin, and Figure 8 shows the slider that is used for the elicitation of preferences concerning the size of the public sector (taken in treatment *EN*).

Amount of correct additions in this round so far: 1 Round 1 of 4 Time: 06:53

Amount of incorrect additions in this round so far: 1

Remaining problems that can be solved in this round: 28 Total wage earned in this round: 429 (minus 171.6)

If your next answer is correct you will receive 390 points (from which 156 will be deducted).

38	39	24	30	58	67	53	60	53	58	83	57	25	13	51	34	66	42	74	63
73	20	12	79	56	32	29	24	79	40	12	69	29	28	19	37	53	77	57	74
24	30	43	25	21	59	25	73	78	41	22	22	58	46	75	82	66	82	13	70
67	23	10	77	35	30	15	25	40	76	14	50	47	58	21	76	56	81	19	30
48	58	20	36	28	35	15	29	49	57	27	62	52	38	59	36	54	42	68	74
33	71	30	53	44	72	65	55	56	49	67	21	78	64	25	67	78	77	46	83
73	62	39	50	78	35	72	22	69	44	20	73	63	13	40	51	45	57	15	45
54	14	71	50	63	42	71	52	70	17	74	83	81	52	46	71	43	34	42	37
25	12	13	23	62	68	71	41	65	43	37	62	31	34	13	23	20	78	36	62
42	16	24	74	38	68	32	56	65	74	54	75	55	67	65	40	59	34	23	28

Largest number in the left matrix plus largest number in the right matrix:

Go to fixed payment option

Figure 6: Screenshot during a work round

How would you describe your mood at the moment? Please choose according to the shown graph.

The form displays a horizontal row of nine square boxes, each containing a stylized face with a number from 1 to 9. The faces show a progression of expressions from neutral (1) to smiling (9). Below the boxes is a legend with radio buttons for each number, and an 'OK' button at the bottom right.

Number	Face Expression
1	Neutral
2	Slightly smiling
3	Smiling
4	Smiling
5	Smiling
6	Smiling
7	Smiling
8	Smiling
9	Smiling

Legend:

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

OK

Figure 7: Screenshot during the self-assessment with the SAM-V-9

Please choose your most preferred position of the slider.

Multiplication factor of the common fund: 3.00

Contribution to the common fund, expressed as a percentage of your wage: 0.00



Figure 8: Screenshot of the slider; public sector size preference

C Appendix (for Online Publication): Additional Data Analysis

C.1 Summary Statistics and Non-Parametric Tests

Table C.1 shows the composition of the sample in the different treatments of the experiment. The means and standard errors of all outcome variables in all four treatments can be seen in Tables C.2 to C.5. Tables C.2 to C.5 also show the p-values of two-sided Wilcoxon rank sum tests (where the null-hypothesis is always no treatment difference between a payroll tax treatment and the corresponding income tax treatment).

Table C.1: Sample composition

	<i>EN</i>	<i>IN</i>	<i>EP</i>	<i>IP</i>
Average age	21.6	22.8	21.7	22.0
Gender male	32	25	24	31
Studies econ + science	36	34	34	30
Lab experience	48	44	44	40

Notes: The table shows the composition of the sample. Each of the controls in the regressions in Section 5 is considered. The numbers reported are for employees only (the total number of employees in each treatment is 50). The age shown is the average age in the treatment, gender is the number of men, studies is the number of subjects studying economics or science, lab experience shows the number of subjects that had participated in economic experiments before.

Table C.2: Non-parametric analysis of public sector size preference (*ps*)

	Employer payroll tax mean (std. error)	Income tax mean (std. error)	Treatment differences p-value Wilcoxon
Nothing in return	51.18 (4.22)	33.00 (2.77)	0.002
Public good	41.18 (3.69)	33.12 (2.96)	0.080

Notes: The table shows the results for individuals' preferences for the size of the public sector (the variable *ps*). Individual outcomes are integers between 0 and 100, larger numbers representing preference for a larger public sector. The p-values stem from two-sided Wilcoxon rank-sum tests.

Our conclusions from the non-parametric analysis are the same as the conclusions from the regressions presented in the main text. We summarize them here briefly. Subjects prefer a larger public sector under a payroll tax levied on employers than under an income tax. This holds whether the tax proceeds are wasted or invested in a public good

Table C.3: Non-parametric analysis of subjective well-being (*sw*)

	Employer payroll tax mean (std. error)	Income tax mean (std. error)	Treatment differences p-value Wilcoxon
Nothing in return	21.84 (0.66)	18.94 (0.84)	0.007
Public good	21.16 (0.77)	21.82 (0.75)	0.594

Notes: The table shows the results of individuals' subjective well-being (the variable *sw*). Individual outcomes are integers between 4 and 36, larger numbers representing higher subjective well-being. The p-values stem from two-sided Wilcoxon rank-sum tests.

Table C.4: Non-parametric analysis of labor supply at the extensive margin

	Employer payroll tax mean (std. error)	Income tax mean (std. error)	Treatment differences p-value Wilcoxon
Measure 1 (<i>le1</i>)			
Nothing in return	1416.3 (34.8)	1525.4 (46.1)	0.133
Public good	1408.8 (39.8)	1462.7 (40.8)	0.268
Measure 2 (<i>le2</i>)			
Nothing in return	1273.9 (45.3)	1347.0 (49.7)	0.167
Public good	1229.2 (39.2)	1328.0 (38.6)	0.035

Notes: The table shows the results of individuals' labor supply at the extensive margin (variables *le1* and *le2*). Higher numbers represent higher labor supply. The p-values stem from two-sided Wilcoxon rank-sum tests.

(the effect seems to be a bit stronger when the tax proceeds are wasted). Subjective well-being is higher under an employer payroll tax than under an income tax if the tax proceeds yield nothing in return. There is a strong interaction, though, between the type of tax and the use of tax proceeds. When the tax proceeds are invested in a public good returned to the employees, the difference in subjective well-being between the different tax types disappears (in line with the mechanisms discussed in Section 4). Labor supply at the extensive margin is higher under an income tax than under a payroll tax.²⁹ The data on labor supply at the intensive margin and job performance are mainly

²⁹The non-parametric tests we employ here have a bit lower statistical power than the regressions in the main text. Therefore it is not surprising that only one of the four tests on labor supply at the extensive margin is significant at the 5%-level. Given that labor supply is higher in all four cases (two measures of labor supply, both with and without public good), and that the other three p-values are relatively low, we interpret the data as overall providing evidence for higher labor supply at the extensive margin under an income tax.

Table C.5: Non-parametric analysis of labor supply at the intensive margin and job performance

	Employer payroll tax mean (std. error)	Income tax mean (std. error)	Treatment differences p-value Wilcoxon
LS intensive margin (<i>li</i>)			
Nothing in return	1570.5 (58.1)	1522.5 (66.2)	0.703
Public good	1426.5 (66.7)	1570.5 (60.6)	0.072
Job performance (<i>jp</i>)			
Nothing in return	20.56 (1.07)	19.64 (1.17)	0.679
Public good	18.40 (1.21)	21.02 (1.23)	0.139

Notes: The table shows the results of individuals' labor supply at the intensive margin and job performance (variables *li* and *jp*). Higher numbers represent higher labor supply or performance. The p-values stem from two-sided Wilcoxon rank-sum tests.

inconclusive, with a hint at larger labor supply and job performance under an income tax when the tax proceeds are not lost.

C.2 An Explorative Analysis of Gender Effects

In the world in- and outside the laboratory, gender differences exist regarding many economic variables (Croson and Gneezy, 2009; for labor supply see e.g. Evers et al., 2008). We therefore explore whether such gender differences exist in our data with respect to the way in which tax framing affects the variables we consider. Note that we are not interested in differences in levels here (e.g., 'do women produce more output?'), but in differences in the treatment effects (e.g., 'is female job performance affected differently by tax framing than male job performance?').

We thus explore whether the results from Section 5 (or similarly Appendix C.1) are driven by either gender. For this (explorative) analysis of our experimental data, regressions are not optimal, because they would flow over with interaction terms (not only the interaction between gender and tax type would have to be included, but also the interactions between gender and the interaction between tax type and use of tax proceeds as well as the interaction between gender and use of tax proceeds, on top of the 'regular' interaction between tax type and use of tax proceeds). Therefore, we look at the data in a similar way as in Appendix C.1, splitting the data into nothing in return and public good treatment data. We then split according to gender and compare the treatment differences for men and women separately.

Table C.6 shows the means of the different outcome variables, split up according to treatment and gender. It is surprising that some of the effects reported above are driven solely by male participants while some others are driven solely by female participants.³⁰

Table C.6: Results split up according to gender

	Employer payroll tax mean m/f	Income tax mean m/f	Treatment differences p-value Wilcoxon m/f
<i>ps</i>			
Nothing in return	54.44/45.39	31.84/34.16	0.009/0.146
Public good	40.46/41.85	36.68/27.32	0.634/0.024
<i>sw</i>			
Nothing in return	21.66/22.17	20.71/17.24	0.727/0.004
Public good	22.42/20.00	21.67/22.05	0.814/0.351
<i>lse1</i>			
Nothing in return	1459.1/1342.5	1588.5/1448.3	0.105/0.510
Public good	1463.1/1361.3	1507.2/1391.0	0.532/0.751
<i>lse2</i>			
Nothing in return	1246.0/1323.4	1406.6/1287.4	0.007/0.726
Public good	1230.3/1228.3	1373.9/1253.1	0.020/0.788
<i>li</i>			
Nothing in return	1534.5/1634.4	1626.7/1418.3	0.353/0.115
Public good	1336.4/1509.8	1627.7/1480.1	0.088/0.333
<i>jp</i>			
Nothing in return	21.06/19.67	22.48/16.80	0.473/0.312
Public good	18.04/18.73	22.53/18.63	0.099/0.881

Notes: The table shows the means of the different outcome variables split according to gender. Larger numbers represent preference for a larger public sector, higher subjective well-being, supply of more labor, and higher performance, respectively. The p-values stem from two-sided Wilcoxon rank-sum tests, comparing treatment effects between employer payroll tax and income tax treatments separately for men and women.

A first thing to observe is that the aggregate results concerning public sector size preferences are driven by men when there is no public good while they are driven by women when there is a public good. After having previously experienced a regime where tax proceeds are lost, men prefer a significantly larger public sector if the tax (that is then

³⁰Of course, the sample sizes are now much smaller. There are 32 (18) men (women) in treatment *EN*, 25 (25) in *IN*, 24 (26) in *EP*, and 31 (19) in *IP*. The two observations that have missing data due to computer problems during the experiment are both from male participants (see Footnote 22). The composition of the sample with a bit more men than women has the consequence that there is slightly higher statistical power to detect treatment differences for men than there is for women.

used to produce a public good) is being levied on the employer. The similar effect for women is smaller. In contrast, when previous experience was with a regime where taxes were already used to produce a public good, men barely respond to differences in who is paying, while women want a much larger public sector if the employer pays the taxes. In terms of subjective well-being, the aggregate result that when proceeds are wasted, people feel better if the tax was paid by the employer, is driven by women. The well-being reported by men barely responds to the framing of the taxes. The labor supply and job performance effects are all mainly driven by men, with highly significant differences (p-values of 0.007 and 0.02) for the second measure of labor supply at the extensive margin (with and without public good, respectively) and marginally significant differences for labor supply at the intensive margin and job performance in the treatments with public good. When it comes down to such labor market responses, women are barely affected by the framing of the tax.

While the results on public sector size preferences are somewhat inconclusive concerning gender effects, the results on subjective well-being and labor supply suggest that gross wage illusion is stronger for men while tax aversion is stronger for women. Most of the economic literature looking at gender effects seems to suggest that women are more sensitive to framing than men (see Croson and Gneezy, 2009). In contrast, our results suggest that the mechanisms underlying gender effects in framing differ between men and women. As a consequence, women are more sensitive to framing for some measures while men are more sensitive to framing for others. In this line of reasoning, women are more sensitive concerning subjective well-being where tax aversion is the dominant mechanism. On the other hand, men are more sensitive to framing concerning the variables traditionally of most interest to economists – labor supply and job performance where gross wage illusion is more important.