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

Short title: The Non-Equivalence of Labour Market Taxes

Matthias Weber† Arthur Schram‡

November 25, 2015

Abstract

Under full rationality, a labour 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. In a real-effort experiment, we study the effects of these 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. Subjective well-being is higher under employer-side taxes while labor supply is lower, at least at the extensive margin. We discuss three mechanisms that may underlie these results.

*Financial support by The Netherlands' Organisation for Scientific Research (NWO grant 406-11-022) and the Research Priority Area Behavioural 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, an anonymous referee, 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, Barcelona, and Mannheim. Most of the work on this article was conducted while Weber was at the University of Amsterdam (CREED). Disclaimer: The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Bank of Lithuania or the Eurosystem. Corresponding author: Matthias Weber, CEFER – Bank of Lithuania & Vilnius University, Totoriu g. 4, 01121 Vilnius, Lithuania.

†CEFER, Bank of Lithuania & Faculty of Economics, Vilnius University. Email: mweber@lb.lt.

‡CREED, University of Amsterdam & Robert Schumann Center for Advanced Studies, EUI (Florence). Email: schram@uva.nl.

doi: 10.1111/ecoj.12365

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 behaviour 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 fulfil this role and LSE is no longer obvious.[1]

In this paper, we study the perception and the behavioural responses to distinct labour taxes that are equivalent in the traditional sense. This is important since labour taxes play a major role in all modern economies and many tax policies are still based on the lessons obtained in traditional public finance.[2] Labour tax equivalence follows from a rational perception of distinct taxes in combination with utility maximizing choices and

[1] Another assumption commonly made in the traditional public finance literature is that individuals have self-regarding preferences. Numerous papers in behavioural 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 solely 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.

[2] A small part of the traditional public finance literature allows for failure of LSE in the labour market due to market frictions. We are not interested here in studying labour 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.

market forces. We specifically study the first element. Rational perception (which in our case is the absence of framing effects) 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 aspects of tax equivalence. Such other aspects have been largely neglected. With boundedly rational agents, 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. Equivalence could also be violated without any direct effect on market prices or quantities if the distinct taxes lead to different levels of subjective well-being.

There are many ways in which labour 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 favour duties levied on the employees' side while left-wing politicians tend to favour 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.[3] People simply might be happier under some tax regimes than under others. In the end, a third effect of labour market taxes is probably to many economists the most obvious. This is their effect on labour 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

[3]For 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).

(labour 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.

Two ways of taxing labour prevail around the world. One is an income tax levied on employees, the other a payroll tax levied on employers.[4] Note that these two types of taxes exist side by side in many countries. 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.

More specifically, we consider the effects of differential perception of theoretically equivalent labour market taxes on each of the three dimensions mentioned above: political-economic preferences, subjective well-being, and labour 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 labour 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

[4]Employer 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 labour market and the term 'employer payroll tax' for a tax (or contribution) on the employers' (demand) side.

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 favourable 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 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 labour 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

[5]Nevertheless, 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).

income (that may subsequently be taxed). Our distinction between an environment where proceeds are lost and one when a public good is produced 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 – net wage illusion, tax loss effect, and warm glow – that would explain such framing effects and discuss how these mechanisms interact to predict the experimental results we observe.

Our results suggest that differences in the way the two taxes are perceived affect behaviour in each of the dimensions that we distinguish between. More specifically, in our experiment employer-side taxes lead to (1) workers preferring a larger public sector; (2) higher subjective well-being of the workers; and (3) lower labour 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.

The organisation of this paper is as follows. Section 1 reviews the related literature and Section 2 presents the experimental design and procedures. This is followed by Section 3, which contains the hypotheses to be tested and the psychological mechanisms we expect to be at work. Section 4 presents the results. In Section 5 we discuss the results and their implications.

1 Related Literature

To our knowledge there are no previous studies investigating the effects of the liability side of a labour tax on political-economic preferences. Nor do we know of any study of the effects of such taxes on subjective well-being. There are, however, other studies examining labour 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 labour 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 that this might be due to differential effects of these tax changes on labour 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. 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 labour supply decisions, neither on the intensive nor on the extensive margin.

There is a (limited) experimental literature on the effects of taxes on labour supply. We know of no such study implementing an employer-employee relationship in the laboratory. Hayashi *et al.* (2013) 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 labour supply differs across treatments. Their most robust finding is that subjects choose to supply less labour in the wage subsidy treatment than in the others. Fochmann *et al.* (2013) investigate whether the gross wage has an influence on the labour 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 'net wage illusion'.[7] Djanali and

[6] There are different approaches to modeling the labour market and thus the impact of labour market taxes. Most prominent are the competitive labour 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. Non-equivalence can arise in exceptional cases via market frictions, however (e.g. Picard and Toulemonde, 2001). None of these approaches allow for non-equivalence due to tax perception.

[7] Fochmann and Weimann (2013) elaborate on Fochmann *et al.* (2013) and explain these findings by tax salience.

Sheehan-Connor (2012) find that experimental subjects work harder for the same net wage when the gross wage is higher and the tax proceeds go to a non-profit organisation, in line with various social preference theories. 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 labour 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 and Kerschbamer and Kirchsteiger (2000) consider taxing in the ultimatum game. The former paper presents results in support of LSE, while the latter presents evidence that LSE does not hold. Finally, Kachelmeier *et al.* (1994), Borck *et al.* (2002), Ruffle (2005), and Cox *et al.* (2012) all examine tax incidence in double auction or posted offer markets. The results are mixed, with the first three papers finding (by and large) support for LSE, while Cox *et al.* (2012) report that LSE does not hold in their data. All in all, the experimental work on LSE in various environments provides inconclusive results.

Considering tax perception more generally, many studies report seemingly irrational behaviour by laboratory participants. An excellent survey is presented in Fochmann *et al.* (2010). 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 labour 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. These 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' behaviour then becomes less elastic to the level of the toll.

2 Experiment

The experiment was conducted in English 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.[8] 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 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 screen-

[8]The software used to run the experiment and the raw data are available online.

shots. 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.[9] When scheduling, we distributed the treatments in a balanced way over mornings and afternoons and across the different days of the week.

2.1 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 problem-solving 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 labour 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 summarises the design. The acronyms *EN*, *IN*, *EP* and *IP* for the four distinct treatments will be used regularly in the remainder.

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. 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

[9] A 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.

Table 1: 2×2 Design

	Employer payroll tax	Income tax
Nothing in return	EN (60)	IN (60)
Public good	EP (60)	IP (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.

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 treatments.[11] As we will explain below, the payment schedule for correctly solved problems of the real effort work task is decreasing. In the public good treatments subjects receive (untaxed) additional earnings from the public good (i.e. from the performance by others) after the experiment, but they do not receive any information on others' performances while taking their decisions. To illustrate tax equivalence between treatments, Table 2 shows an example of wages, taxes, and net earnings in the experiment. In this example it is assumed that one employee solves only the first problem correctly (the tax schemes are equivalent as illustrated, no matter how many problems are solved correctly).

[10] As a consequence, a tax revenue of r points in a group with 5 employees yields $\frac{1.3r}{5}$ 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. Though the public good treatments involve (mandatory) contributions, this experiment is ill-suited to isolate free-riding motives because many factors are involved in the decision to work (and thus contribute via taxes to the public good). Our results will show that labour supply and job performance are not higher in the public good treatments than in the corresponding nothing in return treatments which suggests that subjects are not focusing on payments to others from the public good when they decide about labour supply and effort. Indirectly, this could be interpreted as some evidence for free riding.

[11] The returns from the public good can be split in a part that is due to own tax payments and a part that is due to the taxes paid by the other employees in the group. Consider a task an employee faces. Denote now by W_P and W_N the *net* wages for this task in the two public good and the two nothing in return treatments, respectively. Denote by $\gamma \cdot W_P$ the tax paid, which is a (mandatory) contribution to the public good in the public good treatments (γ thus corresponds to the tax rate in the employer payroll tax treatments but not in the income tax treatments, because there the tax base is not the net wage). In the nothing in return treatments the employee receives W_N, which compares to $W_P + \frac{\gamma \cdot W_P \cdot 1.3}{5}$ in the public good treatments. W_P and W_N are chosen to equate these two returns.

Table 2: Illustration of Wages, Taxes, and Earnings in the Experiment

	EN	*IN*	*EP*	*IP*
Gross wage	280.8	468	239.32	398.87
Income tax (40%)		*-187.2*		*-159.55*
Net wage	280.8	280.8	239.32	239.32
Own performance PG benefits			41.48	41.48
Net earnings employee	**280.8**	**280.8**	**280.8**	**280.8**
Employer tax (66.7% of gross wage)	*-187.2*		*-159.55*	
Total labour costs (wage + tax)	468	468	398.87	398.87
Net earnings employer	49.8	49.8	49.8	49.8

Notes: This table illustrates the equivalence of the taxes in the different treatments for the example that an employee has solved the first problem correctly. The numbers are in points.

2.2 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 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 in order to avoid (unmeasured) preconceptions that some subjects might have with respect to terms referring directly to taxes. Figure 1 gives a schematic overview of the timeline; the different parts will be explained in the following.

2.2.1 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 2 shows a screenshot from the work task (taken from treatment *IN*; see Appendix B for a larger version). 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.

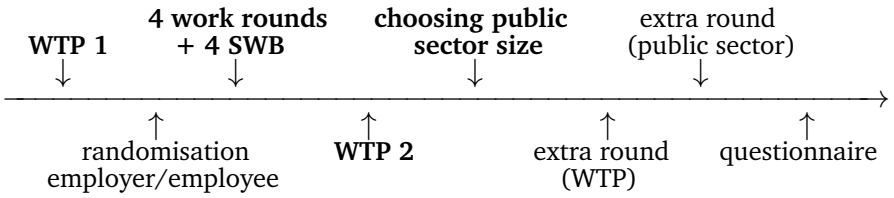

Figure 1: Timeline of the Experiment

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 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.

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.[12] This limit and the way the random numbers are generated make guessing a very unsuccessful strategy. 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. In the employer payroll tax treatments, subjects do not see the tax the employer has to pay. Subjects can also see the number of correct and incorrect additions so far and the remaining time. In the public good treatments, this screen is the same as in the corresponding nothing-in-return treatments.

This procedure is repeated in four independent and identical rounds. The total number of correctly solved problems is a measure of job performance. During these rounds,

[12]The mean, median, minimum, and maximum numbers of correctly solved problems in a single round (across all treatments) are 4.97, 5, 0, and 12, respectively. Note, however, that subjects have the option to stop working before a round ends by switching to a fixed payment option, which will be described below.

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 (Charness *et al.*, 2007; Sutter, 2009).

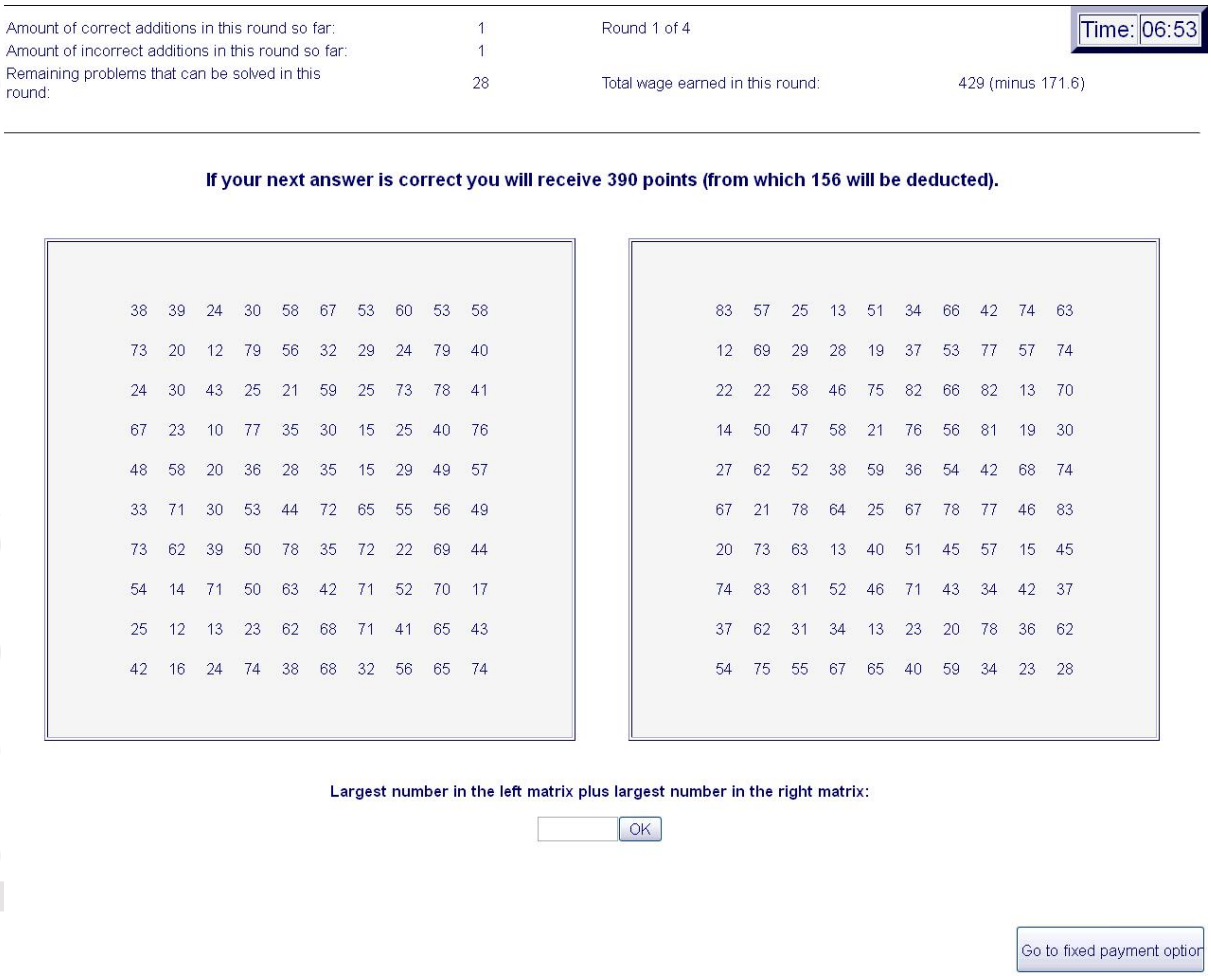

Figure 2: Screenshot during a Work Round (Taken in Treatment *IN*)

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 (cf. Table 2). With each attempt (whether correct or not) the payment for the next correct addition

decreases by 23.4 points. 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).[13]

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 (the screen shows only the remaining time and informs subjects on their earnings from the fixed payment option). 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 labour supply at the intensive margin.

2.2.2 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. In the public good treatments, subjects are also reminded on this screen that they will receive additional earnings from the public good at the end of the experiment. 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

[13] Decreasing payments make it more likely that subjects use a fixed payment outside option (which will be described below), 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 *EN* (net=gross). The gross wage in *IN* is $\pi = max(468 - 39x, 0)$, which leads, with a tax rate of 40%, to the same net wage as in *EN*. The gross wages in *EP* (gross=net) and *IP* are $\pi = max(239.32 - 19.94x, 0)$ and $\pi = max(398.87 - 33.24x, 0)$, respectively. Both of these gross wages lead to the same net wages. They furthermore lead to earnings due to own performance that is equal to the net wage in *EP* and *IP* (see Footnote 11). In the public good treatments subjects receive additional earnings from the tax payments by others, which they cannot influence. Over the four regular work rounds these extra earnings lie between 0 (if no other group member solves any problem correctly) and 4314 points (if all co-workers solve all problems correctly).

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 3.

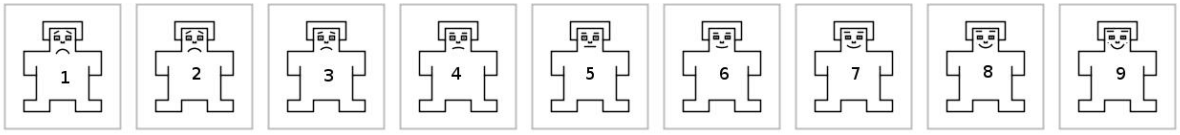

Figure 3: Subjective Well-Being Self-Assessment Manikin

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

2.2.3 Measuring Labour 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.[14] 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.[15] 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.

[14]For subjects subsequently randomised to be employers the revealed willingness to pay has no further consequence.

[15]This means that the extra round takes place with probability one half. The random draw is always the same for all subjects of a group.

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 (if it lies within the price interval, which should be the case for the vast majority of subjects). 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. Employers receive earnings in the extra round in the same way as in the regular rounds. The stated willingness to pay provides us with a (first) measure of labour supply at the extensive margin.[16]

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 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 labour 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).

2.2.4 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

[16] In reality, people usually do not start participating in the labour market by just clicking on a button somewhere. There are 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 labour supply at the extensive margin than a mere statement whether one wants to participate or not.

treatments (*EN* and *IN*), a public good is introduced, such that the taxes are no longer lost, just as in the public good treatments. In all four treatments, the tax rate and the multiplication factor for the public good are now 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 4. 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.

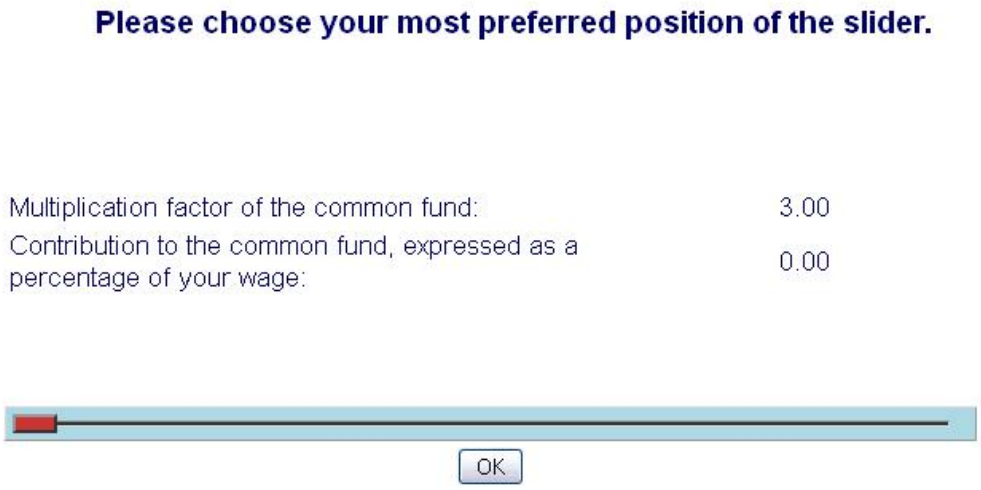

Figure 4: Slider for the Elicitation of Public Sector Size Preference

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.

The increase in the tax rate is different for the income tax and employer payroll tax treatments, because the tax bases are different. Instead of equalizing the tax rate for each slider position across treatments, tax rates on the slider are determined in a way that each position of the slider leads (in the absence of behavioural responses to the tax framing) to the same tax revenue. The 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 corresponds to 75% of the employer's total labour cost at the right end; the tax is

always expressed as a percentage of the employee's (gross) wage.[17] In all treatments, neither the employer's gross revenue nor her net earnings per correctly solved problem change with the slider position. Thus, in the employer payroll tax treatments, the wage (which is the tax base) has to adjust with the slider position in order to obtain tax revenue equivalence between the treatments. That gross revenues and net profits of the employer stay the same and that therefore the wage has to decrease if a higher tax is chosen in *EN* and *EP* is very salient for subjects (they are told and questioned about this right before the task). However, the exact adjustment of the wage is not shown to subjects when they move the slider.

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. The number corresponding to the chosen slider position provides us with a measure of the subject's preferred size of the public sector. Note that if there is an anchoring effect where subjects choose the tax rates that they experienced in the earlier work rounds (40% in the income tax treatments and 66.7% (of a lower tax base) in the employer payroll tax treatments), this would lead to the same slider position in all treatments.

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 tax-rate space.

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 income tax treatments are still absolutely equivalent, whereas the nothing in return and the public good treatments are now somewhat 'less equivalent'. Subjects

[17] 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.

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*.

Table 3 illustrates the equivalence between the treatments *EP* and *IP* with a simple example. In this example we compare the implications of the choice of slider position 20. We assume (as in Table 2) that one subject has solved the first problem correctly. The equivalence for other slider positions or numbers of correct solutions and that between treatments *EN* and *IN* can be illustrated analogously.

Table 3: Illustration of the Equivalence of the Same Slider Positions in *EP* and *IP*

	Slider pos.	Mult. factor	Tax rate	Tax base / gross wage	Net wage	Tax revenue	Net earnings employer
EP	20	2.55	17.65%	339.04	339.04	59.83	49.8
IP	20	2.55	15%	398.87	339.04	59.83	49.8

Notes: This table shows the equivalence between the same slider positions in EP and IP (based on a chosen slider position of 20 and assuming that one subject solves the first problem correctly). Subjects see the multiplication factor and the tax rate when they move the slider (as in this table, rounded to the second decimal digit). The tax revenue is the contribution to the public good, paid by the employer in EP and by the employee in 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).[18] 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) is revealed. Before being paid, participants are asked to fill out a short questionnaire, including questions concerning gender, age, field

[18]Before 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 fact that we are not dealing with risky choices (for which the test was 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.

of study and experience in laboratory experiments (see Appendix A).

3 Hypotheses and Mechanisms

The main goal of this paper is to test the effects of the liability side of a labour tax – represented by pure framing – on outcomes of the three dimensions we discussed. We use null-hypotheses about treatment effects arising from rational choice behaviour. These imply no differences across treatments.[19] However, note that the two taxes are also equivalent assuming various other outcome based preferences, such as 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.

Now, we discuss which psychological mechanisms we expect may govern participants' behaviour. These mechanisms arise partly from salience considerations (Chetty *et al.*, 2009; Finkelstein, 2009; Fochmann and Weimann, 2013; Rupert and Wright, 1998; Sausgruber and Tyran, 2005, 2011), but they are not exclusively based on tax salience. The mechanisms 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.* (2013) call 'net wage illusion' (as observed by Fochmann *et al.*, 2013, and Hayashi *et al.*, 2013, in different settings).

[19] As 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 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).

Strictly speaking, there are no rational choice predictions for subjective well-being, because decisions are not incentivised. We also use the null-hypotheses of no difference in outcomes here.

(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 call this 'tax loss effect'.

(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.

Each of these mechanisms has different implications under an employer payroll tax than under an income tax. Before we discuss in more detail how these mechanisms interact and influence participants' behaviour in the decisions and tasks we are investigating, we first briefly describe the consequences of each mechanism in more general terms. Net wage illusion leads individuals to overestimate their net earnings and therefore makes them value work (including payment) higher under an income tax than under a payroll tax. The tax loss effect makes individuals value work (including payment) less under an income tax than under an employer payroll tax, because they dislike the fact that they have to bear the taxes. Warm glow only exists when taxes are used as contributions to the production of public goods and leads to individuals valuing work (including payment) more when the tax is levied on the employee's side, as they then have the feeling of doing a good deed. We allow each of these mechanisms to have an impact that differs across the decisions and tasks participants face.

We start with preferences regarding the size of the public sector. Net 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).[20] The tax loss effect 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 labour market. Hence, a larger public sector is preferred in the payroll tax treatments. The warm glow effect goes in the same direction as net wage illusion. Individuals subject to warm glow may prefer a larger public sector when they are funding it. It is important to note here, that there is always a public good in the extra round used to measure preferences for the size of the public sector. Hence, there is no distinction between the nothing-in-return and public good treatments. Thus, the warm glow effect could apply to all cases. With two effects pointing in the direction of a higher preferred size in the income tax treatment and one effect predicting the opposite, the aggregate effect could go either way.[21]

Next, consider subjective well-being. Net 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. The tax loss effect 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, warm glow may also play a role. This effect aligns with the effect of net 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 labour supply and job performance, we again need to consider the aggregate of the three effects. Net wage illusion directly implies that people will supply more labour (at both the intensive and extensive margins) and perform better in the income tax treatments.[22] The tax loss effect affects labour supply decisions in the opposite

[20]The normality of public goods is a common assumption (e.g., Sugden, 1984, Morton, 1987, or Andreoni, 1990). For indirect empirical support, see, e.g., Schram (1990).

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

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

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 labour supply and better performance in the income tax treatment because of the perceived extra benefits related to helping others. Once again, net wage illusion and warm glow point in one direction, and the tax loss effect points to the opposite.

We summarise the effects of the three mechanisms on our outcome variables in Table 4. It turns out that for all of the variables we consider, net wage illusion leads to a higher outcome under an income tax than under a payroll tax levied on employers while the tax loss effect leads to the opposite effect. The warm glow effect aligns with the effect of net wage illusion (in the treatments where it occurs).

Table 4: Effects of the Mechanisms on the Outcomes

	Treatment condition	Public sector size pref.	Subjective well-being	Labour supply (ext. & int.)	Job performance
Net wage illusion	NIR/PG	$ps^I > ps^E$	$sw^I > sw^E$	$ls^I > ls^I$	$jp^I > jp^E$
Tax loss effect	NIR/PG	$ps^I < ps^E$	$sw^I < sw^E$	$ls^I < ls^I$	$jp^I < jp^E$
Warm glow	NIR	$ps^I > ps^E$	—	—	—
	PG	$ps^I > ps^E$	$sw^I > sw^E$	$ls^I > ls^I$	$jp^I > jp^E$

Notes: This table shows the effects that the behavioural mechanisms have on the outcomes in the income tax treatments (denoted by ^{I}) compared to the employer payroll tax treaments (denoted by ^{E}). For warm glow we indicate the effects of the mechanisms separately for the nothing-in-return (NIR) and the public good (PG) treatments. We denote by ps the public sector size preference; by sw the subjective well-being; by ls the labour supply (at the extensive or intensive margin); and by jp the job performance.

Aside from the fact that we expect these mechanisms to lead to the rejection of rational behaviour 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, labour supply, or job performance only exists in the public good treatments. This can be seen in Table 4. This further shows that for subjective well-being, labour supply, and job performance alike, warm glow leads to a higher outcome under an

income tax than under an employer payroll tax. Thus, if our mechanisms drive the results and warm glow plays a role, there should be a positive interaction effect between the income tax treatment condition and the public good treatment condition for subjective well-being, labour supply, and job performance. We will test this, below. As noted above, our mechanisms do not predict an interaction effect between income tax and public good treatment for public sector size preference, because a public good is always used in the corresponding extra round. Without a public good it would always be optimal to chose a zero tax rate.

4 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 labour supply at the extensive margin, where we have 60 observations per treatment (as this was measured before the randomisation into employees and employers).[23] Because subjects receive no feedback about others' decisions until the end of the experiment, we can treat observations across individuals as statistically independent. The outcome variables we consider are preferences for the size of the public sector (as measured by the chosen slider position), subjective well-being (as measured by the sum of the choices with the self-assessment manikin), both measures of labour supply at the extensive margin (as given by choices in the two BDM tasks), labour supply at the intensive margin (as measured by the total number of seconds in the work mode over the four regular rounds), and job performance (as measured by the total number of problems solved correctly over the four regular rounds).

We present both non-parametric tests and censored regression models in the analysis of our results. Regressions allow us to observe the effects of the interaction between the

[23] Due to computer problems, one observation in *IN* has missing data on subjective well-being and one observation in *IP* has missing data on subjective well-being, intensive labour supply, and job performance.

tax type and the use of the tax proceeds and to control for observable characteristics. With respect to the latter, we use age, gender (1 if male), field of study (1 if studying at the faculty of economics and business or at the faculty of science), and laboratory experience (1 if having previously participated in at least one laboratory experiment).[24] As is common in experimental papers, we focus on the directions of treatment effects rather than on levels and quantifications (the meaning of exact quantities in laboratory experiments is often limited).

4.1 Public Sector Size Preference

Table 5 shows descriptive statistics of public sector size preferences (i.e. slider positions) and p-values from two-sided Wilcoxon rank-sum tests. Table 6 shows the coefficient estimates from censored regressions (both with and without covariates).[25]

Table 5: Overview and Rank-Sum Tests, Public Sector Size Preference

	Employer payroll tax mean (std. error)	Income tax mean (std. error)	Treatment diff. 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: 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. The p-values stem from two-sided Wilcoxon rank-sum tests.

Both, when taxes are lost and when proceeds are invested in a public good, subjects

[24] Appendix C.2 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.

[25] The optimal slider position cannot easily be determined in the experiment as it depends on subjects' reactions to the tax (and to the use of the tax proceeds and to the tax framing). If labour supply/performance were completely inelastic to the tax the optimal slider position would be 45 (which corresponds to an income tax of 33.75%). Given that a substantial part of the tax proceeds is paid back to an employee via the public good (around 40% of the paid tax at the position 45, more if it is below) and given the decreasing marginal pay for completing the work task, it may be that a true optimum in the experiment is not far below 45. However, results on whether chosen slider positions are above or below such an optimum do not arise from treatment effects. Because the comparative statics of treatment comparisons are generally considered to be most relevant for behaviour outside of the laboratory, it is questionable whether the relative position to the optimum has much meaning for the world outside of the laboratory.

prefer a larger public sector when the tax is levied on the employer's side. These differences are highly significant in the case where tax proceeds are lost and marginally significant between the public good treatments. The results from the regression analysis strengthen this result – the coefficient of the tax condition dummy is negative (and significant at the 1%-level, independently of whether or not covariates are included in the regression). The interaction coefficient is not significant, which is not particularly surprising, given that the mechanisms we outlined in Section 3 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 is even significant at the 5%-level if no covariates are included in the regression). This 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.

Table 6: Regression Results for Public Sector Size Preference

	Coefficient	Std. error	Coefficient	Std. error
Tax (1=income tax)	-20.10***	(5.48)	-17.15***	(5.54)
PG (1=public good)	-11.50**	(5.49)	-9.91*	(5.46)
Tax * PG	11.62	(7.73)	9.04	(7.77)
Intercept	52.31***	(3.89)	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)
Observations	200		200	

Notes: Individual outcomes are integers between 0 and 100. 13 observations are left censored and 11 right censored. *,**, and *** stand for significance at the 10-, 5-, and 1-percent level.

Recalling that of the three mechanisms we discussed, net wage illusion and warm glow both predict that a larger public sector is preferred in the income tax treatment, our findings suggest that the tax loss effect plays a dominant role when individuals determine their preference for the size of the public sector.

4.2 Subjective Well-Being

Table 7 gives an overview of the results on subjective well-being. Table 8 shows the corresponding regression results.[26]

Table 7: Overview and Rank-Sum Tests, Subjective Well-Being

	Employer payroll tax mean (std. error)	Income tax mean (std. error)	Treatment diff. 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: Individual outcomes are integers between 4 and 36, measured by the sum of the values of all four self-assessments. Larger numbers represent higher subjective well-being. The p-values stem from two-sided Wilcoxon rank-sum tests.

We can see in Table 7 that participants report higher subjective well-being under an employer payroll tax when the tax proceeds are lost. This finding is significant at the 1%-level. Subjective well-being is not affected by the tax treatment when the tax proceeds are invested in a public good. In line with the results from the non-parametric tests, the regression results show a highly significant tax dummy coefficient. Furthermore, 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 shows 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; this explains the outcomes in Table 7).

Looking back at our mechanisms, these results suggest that when tax proceeds yield nothing in return, the tax loss effect outweighs the effect of net 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 where it predicts higher subjective-well being under an income tax (thus a positive interaction term). The finding of the interaction effect thus corroborates our view that

[26]Consideration of the four separate self-reports shows consistent behaviour and no indication of trends.

Table 8: Regression Results for Subjective Well-Being

	Coefficient	Std. error	Coefficient	Std. error
Tax (1=income tax)	-2.90***	(1.06)	-2.80***	(1.08)
PG (1=public good)	-0.71	1.06	-0.62	(1.05)
Tax * PG	3.59**	(1.50)	3.09**	(1.51)
Intercept	21.84***	(0.75)	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)
Observations	198		198	

Notes: Individual outcomes are integers between 4 and 36. 1 observation is left censored, none are right censored. *,**, and *** stand for significance at the 10-, 5-, and 1-percent level.

the mechanisms we described – in particular, warm glow – are at work.

4.3 Labour Supply and Job Performance

The measures on labour supply and job performance we use can be divided into two parts. The first consists of two measures of labour supply at the extensive margin (once extracted before the randomisation and experiencing the work task and once after). These measures capture the willingness to bear a cost in order to work. The extensive labour supply decisions are taken outside of the regular work environment in our experiment, but this is arguably also the case in the world outside of the laboratory.[27]

The second part of our measures concern labour 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 labour by quitting the task for the remainder of the current round. labour supply at the intensive margin

[27]Outside 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. 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).

and job performance are of course related; participants spending more time solving problems are likely to solve more of them. They also differ substantially from the two measures of labour supply at the extensive margin. For this reason, we will separate the discussion and first discuss labour supply at the extensive margin and then labour supply at the intensive margin and job performance.[28]

4.3.1 Labour Supply at the Extensive Margin

Table 9 gives an overview for the outcomes on labour supply at the extensive margin. It shows that independently of which measure one considers and of whether one considers the nothing in return treatments or the public good treatments, labour supply is always higher under an income tax. However, while the differences are always considerable (more than two standard errors in two cases and still more than one standard error in the other two cases), only one of the differences is statistically significant at the 5%-level when tested with a Wilcoxon rank-sum test. Considering Table 9, there seems thus to be some (admittedly rather weak) evidence that extensive labour supply is higher under an income tax than under an employer payroll tax.

Table 9: Overview and Rank-Sum Tests, Labour Supply at the Extensive Margin

	Employer payroll tax mean (std. error)	Income tax mean (std. error)	Treatment diff. p-value Wilcoxon
Measure 1			
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			
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: Individual outcomes are integers between 1056 and 2400. These numbers represent the outcomes of the two measures of labour supply at the extensive margin (the preplay and the postplay willingness to pay to participate in an extra round). Larger numbers represent higher labour supply. The p-values stem from two-sided Wilcoxon rank-sum tests.

[28]The 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.

Tables 10 and 11 show the regression results, separately for both measures. The regression results for the first measure show a positive coefficient of the tax dummy that is significantly different from zero (only marginally so if no covariates are included in the regression). Furthermore, looking at the regression coefficients, one can see that there is a strong and statistically highly significant positive effect of being male. Thus, men are willing to pay significantly more to participate in an extra round than women. Not being able to control for gender may thus explain why the treatment differences for the first measure are not statistically significant when performing non-parametric tests.[29] The coefficient of the interaction term is not significantly different from zero. Moreover, considering the sum of the tax dummy and the interaction term suggests no effect of the tax framing in the public good treatment.

Table 10: Regression Results for Labour Supply at the Extensive Margin, Measure 1

	Coefficient	Std. error	Coefficient	Std. error
Tax (1=income tax)	116.52*	(62.75)	125.77**	(62.39)
PG (1=public good)	-7.64	(62.74)	12.42	(61.95)
Tax * PG	-48.57	(88.59)	-85.91	(87.84)
Intercept	1398.48***	(44.38)	1378.04***	(220.72)
Age			0.02	(9.37)
Gender (1=male)			142.80***	(45.28)
Studies (1=econ+science)			10.90	(49.92)
Lab experience (1=yes)			-84.08	(67.99)
Observations	240		240	

Notes: Individual outcomes are integers between 1056 and 2400. 20 observations are left censored and 4 are right censored. *,**, and *** stand for significance at the 10-, 5-, and 1-percent level.

For the second measure (where statistical power is a bit lower, because it was elicited only for employees), the sign of the income tax coefficient is positive, but not significantly different from zero. The coefficient of the interaction term between the tax and public good treatments is insignificant and close to zero. Note that in this case, an outcome that is statistically significant when conducting a Wilcoxon rank-sum test

[29] While the randomisation generally balances characteristics like gender well, it does not balance them perfectly. In this experiment, the number of male subjects per treatment varies between about one half and about two thirds. More on the sample characteristics can be found in Table C.1 in Appendix C.1.

Table 11: Regression Results for Labour Supply at the Extensive Margin, Measure 2

	Coefficient	Std. error	Coefficient	Std. error
Tax (1=income tax)	87.94	(77.74)	94.51	(79.18)
PG (1=public good)	-44.64	(78.06)	-35.93	(78.16)
Tax * PG	27.17	(109.82)	-3.24	(110.95)
Intercept	1218.32***	(55.38)	1282.63***	(281.23)
Age			-1.09	(11.95)
Gender (1=male)			96.51*	(57.70)
Studies (1=econ+science)			-77.71	(64.51)
Lab experience (1=yes)			-48.44	(85.97)
Observations	200		200	

Notes: Individual outcomes are integers between 1056 and 2400. 42 observations are left censored and 6 are right censored. *,**, and *** stand for significance at the 10-, 5-, and 1-percent level.

(measure 2 between the public good treatments) loses its statistical significance in a regression setting.

Looking at the overall evidence on labour supply at the extensive margin, there is some evidence that this labour supply is higher under an income tax than under an employer payroll tax. The evidence is weaker than for the findings on public sector size preferences and subjective well-being. However, the fact that the measures we extracted are higher in all four cases under an income tax and the fact that a part of these differences is statistically significant with non-parametric tests while another part is significant in a regression analysis where one can control for observed covariates (in particular gender) gives support to the claim of higher labour supply under an income tax. Aggregating the two measures for the statistical analysis does not give a different picture and confirms the conclusion that there is some support that labour supply at the extensive margin is higher under an income tax.[30] As for the three mechanisms, net wage illusion seems to

[30]For the non-parametric analysis, one can construct an aggregate measure for each employee by taking the average of the two measures (after dividing by the mean outcome of each measure). Wilcoxon rank-sum tests then yield that the difference between the nothing-in-return treatments is not significant, while the difference between the public good treatments is marginally significant. A linear regression analysis where the data concerning both measures are pooled and standard errors are clustered at the subject level (including a dummy to capture the general trend) leads to coefficient estimates in line with the results reported in Tables 10 and 11. The income tax dummy is marginally significant without additional covariates and significant with the usual covariates (with strong gender effects). The public good dummy and the interaction coefficient are not significant (without and with covariates). More information is available upon request.

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

4.3.2 Labour Supply at the Intensive Margin and Job Performance

Table 12 gives an overview of the outcomes on labour supply at the intensive margin and job performance. Both measures are larger under a payroll tax when the tax proceeds are wasted, but larger under an income tax when tax proceeds are invested in a profitable public good (the differences in the nothing in return treatments are very small, however). Only one of the findings is (marginally) statistically significant.

Table 12: Overview and Rank-Sum Tests, Labour Supply at the Intensive Margin and Job Performance

	Employer payroll tax mean (std. error)	Income tax mean (std. error)	Treatment diff. p-value Wilcoxon
Intensive margin			
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			
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: Individual outcomes are between 0 and 1920 for labour supply at the intensive margin, measured by the total time spent on the work task in the four regular rounds. Individual outcomes are positive integers for job performance, measured by the sum of correctly solved problems in the four regular rounds (the maximum achieved in the data is 37). Larger numbers represent higher labour supply and better job performance. The p-values stem from two-sided Wilcoxon rank-sum tests.

Tables 13 and 14 show the regression outcomes for labour supply at the intensive margin and job performance. For both outcome variables (and independently of whether or not covariates are included), the coefficient of the income tax dummy is negative, but small and insignificant. For both variables, the coefficient of the interaction of tax and public good condition is positive (it is marginally significant for labour supply at

the intensive margin). Considering the sum of the tax dummy and the interaction term suggests no effect of the tax framing in the public good treatment (or, at best, hints slightly at larger labour supply at the intensive margin and higher job performance). 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 in Table 14 is positive and significantly different from zero.

Table 13: Regression Results for Labour Supply at the Intensive Margin

	Coefficient	Std. error	Coefficient	Std. error
Tax (1=income tax)	-78.59	(108.28)	-86.45	(111.44)
PG (1=public good)	-190.85*	(107.92)	-198.75*	(108.42)
Tax * PG	263.98*	(153.21)	281.49*	(155.95)
Intercept	1658.14***	(77.35)	1636.92***	(390.39)
Age			1.61	(16.81)
Gender (1=male)			-21.96	(80.57)
Studies (1=econ+science)			143.60	(89.72)
Lab experience (1=yes)			-107.49	(121.60)
Observations	199		199	

Notes: Individual outcomes are between 0 and 1920. No observations are left censored, 41 are right censored. *,**, and *** stand for significance at the 10-, 5-, and 1-percent level.

Table 14: Regression Results for Job Performance

	Coefficient	Std. error	Coefficient	Std. error
Tax (1=income tax)	-0.87	(1.67)	-0.53	(1.67)
PG (1=public good)	-2.22	(1.67)	-1.87	(1.64)
Tax * PG	3.55	(2.37)	3.14	(2.34)
Intercept	20.51***	(1.18)	18.91***	(5.88)
Age			-0.04	(0.25)
Gender (1=male)			1.87	(1.21)
Studies (1=econ+science)			3.20**	(1.35)
Lab experience (1=yes)			-1.13	(1.81)
Observations	199		199	

Notes: Individual outcomes are positive integers. 4 observations are left censored. *,**, and *** stand for significance at the 10-, 5-, and 1-percent level.

Examining the evidence on intensive labour supply and job performance, we are mainly

left with a null-result. The mechanisms we discussed seem to be either balanced (net wage illusion and the tax loss effect 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 for both outcome variables and marginally significant for labour supply at the intensive margin is a further (albeit relatively weak) indication that warm glow plays a role in subjects' decisions.

5 Discussion

The question whether a labour tax levied at the employees' side of the labour 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 employer's side, while labour 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, net wage illusion, tax loss effect, 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 labour supply.

The policy implications of our results are non-trivial as they show that who formally pays the tax affects individuals in different ways. 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. This finding thus suggests that, ceteris paribus, it would be better to levy

taxes on the employer's side. Concerning labour supply, most economists would probably agree that taxes should be chosen that minimise the disruption of the price mechanism in the labour market. As labour taxes are generally thought to reduce labour supply, this implies that the tax should be chosen that maximises labour supply. If people work more under one tax scheme than under the other, using this scheme brings society closer to first best and thus increases social welfare. The experimental findings concerning labour supply are that people tend to supply more labour under an income tax than under an employer payroll tax (at the extensive margin; there are indications of similar effects at the intensive margin and performance, see Fochmann *et al.*, 2013). Thus, the optimal policy for the government when considering only labour supply responses would be to rely only on income taxes. It is less straightforward to draw policy conclusions from our results regarding public sector size preferences. 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 labour taxes is optimal that induces individuals to prefer this size of the public sector. However, while it is already difficult in general to derive the optimal size of the public sector in this experiment, one should not make inferences from this experiment about the distance to the optimum in the real world (see Footnote 25).

The aspects we investigated lead, when considered separately, to distinct implications and it is quite possible that the optimal policy involves a mix of labour 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 minimises collection and compliance costs, even in the absence of labour market frictions. 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 favour.

Our results share some features with those reported by others. Sausgruber and Tyran

(2005, 2011) and our paper have in common that an individual's willingness to pay taxes is larger if these taxes are not levied on this individual directly (thus when individuals irrationally feel that they do not pay the tax). Furthermore, a commonality of these papers is that the liability side of taxes seems to have a much stronger effect on political preferences and voting decisions than on 'hard' economic decisions (such as labour supply or market behaviour in a buyer-seller environment). With Blumkin *et al.* (2012) and Fochmann *et al.* (2013) our paper shares the finding that labour supply tends to be higher when a tax invites to a higher perception of the (real) net wage. Note that the tax itself can be relatively less salient than in the comparison tax regime (Blumkin *et al.*, 2012) or more salient (Fochmann *et al.*, 2013, and our paper).

A potential drawback of laboratory experiments is a limited external validity due to the artificiality of the setting or the subject pool used. There are, of course, no guarantees that subjects' choices in a laboratory experiment fully reflect their behaviour outside of the laboratory; one could easily argue that no experiment, no theory, and no econometric field data analysis fully reflects real-world behaviour. In fact, the assumptions needed to generalise laboratory findings to the field are precisely the same as those needed to establish the generalisability of results obtained by using observational field data (Falk and Heckman, 2009). Moreover, laboratory experimentation provides some advantages that are much more difficult to achieve otherwise. For example, knowledge about causal mechanisms is often crucial for understanding the world outside the laboratory. This requires controlled variation and the laboratory allows for such control in a way that is generally impossible outside of the laboratory.

In addition, with a careful experimental design one can overcome problems of generalizability and gain knowledge about specific features that are relevant for the real world. Problems with external validity can to a large degree be avoided if one focuses on directions of treatment effects rather than on levels and quantifications (see, e.g., Charness and Kuhn, 2011). This is the approach we have taken in this paper. As long as general trade-offs are concerned (e.g. a trade-off between earning money and

experiencing disutility from working, a trade-off that is commonly assumed in labour economics and that we have induced in the laboratory) and as far as only directions of treatment effects are concerned, the results tend to be robust and can be expected to carry over to the real world (see Schram, 2005, and the references therein). In comparison, claims on quantities and levels should be made cautiously – in fact, we discourage the reader from making such claims based on our paper (for example, we refrain from making claims such as that a y-percent increase of some variable could be achieved by a shift in the liability side of the tax). Finally, it goes without saying that our understanding of liability-side (non-)equivalence would benefit from laboratory evidence being confirmed or challenged by further studies involving other methodologies.

Summing up, it seems clear from our results that boundedly rational behaviour 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*, vol. 35(2), pp. 268–284.

Andreoni, J. (1990). 'Impure altruism and donations to public goods: a theory of warm-glow giving', *The Economic Journal*, vol. 100(401), pp. 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 Paper*.

Becker, G.M., Degroot, M.H. and Marschak, J. (1964). 'Measuring utility by a single-response sequential method', *Systems Research and Behavioral Science*, vol. 9, pp. 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*, vol. 56(6), pp. 1200–1219.

Borck, R., Engelmann, D., Müller, W. and Normann, H.T. (2002). 'Tax liability-side equivalence in experimental posted-offer markets', *Southern Economic Journal*, pp. 672–682.

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*, vol. 25(1), pp. 49–59.

Charness, G. and Kuhn, P. (2011). 'Lab labor: What can labor economists learn from the lab?', in (O. Ashenfelter and D. Card, eds.), *Handbook of Labor Economics*, pp. 229–330, vol. 4, Elsevier.

Charness, G., Rigotti, L. and Rustichini, A. (2007). 'Individual behavior and group membership', *American Economic Review*, vol. 97(4), pp. 1340–1352.

Chetty, R., Looney, A. and Kroft, K. (2009). 'Salience and taxation: Theory and evidence', *American Economic Review*, vol. 99(4), pp. 1145–77.

Conlisk, J. (1996). 'Why bounded rationality?', *Journal of Economic Literature*, vol. 34, pp. 669–700.

Cox, J.C., Rider, M. and Sen, A. (2012). 'Tax incidence: Do institutions matter? An experimental study', *Experimental Economics Center Working Paper 2012-17*.

De Bartolome, C.A.M. (1995). 'Which tax rate do people use: Average or marginal?', *Journal of Public Economics*, vol. 56(1), pp. 79–96.

Djanali, I. and Sheehan-Connor, D. (2012). 'Tax affinity hypothesis: Do we really hate paying taxes?', *Journal of Economic Psychology*, vol. 33(4), pp. 758–775.

Evers, M., De Mooij, R. and Van Vuuren, D. (2008). 'The wage elasticity of labour supply: A synthesis of empirical estimates', *De Economist*, vol. 156, pp. 25–43.

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

Fehr, E. and Goette, L. (2007). 'Do workers work more if wages are high? Evidence from a randomized field experiment', *American Economic Review*, vol. 97(1), pp. 298–317.

Fehr, E. and Schmidt, K.M. (1999). 'A theory of fairness, competition, and cooperation', *The Quarterly Journal of Economics*, vol. 114(3), pp. 817–868.

Ferrer-i Carbonell, A. (2013). 'Happiness economics', *SERIEs*, pp. 1–26.

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

Fochmann, M., Kiesewetter, D., Blaufus, K., Hundsdoerfer, J. and Weimann, J. (2010). '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*, vol. 81(1), pp. 230–242.

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

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

Hayashi, A., Nakamura, B.K. and Gamage, D. (2013). 'Experimental evidence of tax salience and the labor-leisure decision: Anchoring, tax aversion, or complexity?', *Public Finance Review*, vol. 41(2), pp. 203–226.

Hossain, T. and List, J.A. (2012). 'The behavioralist visits the factory: Increasing productivity using simple framing manipulations', *Management Science*, vol. 58(12), pp. 2151–2167.

Irtel, H. (2007). 'Pxlab: The psychological experiments laboratory', Version 2.1.11. Mannheim (Germany). http://www.pxlab.de (last accessed April 25, 2011).

Kachelmeier, S.J., Limberg, S.T. and Schadewald, M.S. (1994). 'Experimental evidence of market reactions to new consumption taxes', *Contemporary Accounting Research*, vol. 10(2), pp. 505–545.

Kahneman, D. and Krueger, A.B. (2006). 'Developments in the measurement of subjective well-being', *Journal of Economic Perspectives*, vol. 20(1), pp. 3–24.

Kerschbamer, R. and Kirchsteiger, G. (2000). 'Theoretically robust but empirically invalid? an experimental investigation into tax equivalence', *Economic Theory*, vol. 16, pp. 719–734.

Lang, P.J. (1985). *The cognitive psychophysiology of emotion: Anxiety and anxiety disorders*, Lawrence Erlbaum.

Lehmann, E., Marical, F. and Rioux, L. (2013). 'Labor income responds differently to income-tax and payroll-tax reforms', *Journal of Public Economics*, vol. 99, pp. 66–84.

Morton, R.B. (1987). 'A group majority voting model of public good provision', *Social Choice and Welfare*, vol. 4(2), pp. 117–131.

Muysken, J., Van Veen, T. and De Regt, E. (1999). 'Does a shift in the tax burden create employment?', *Applied Economics*, vol. 31(10), pp. 1195–1205.

Oswald, A.J. (1985). 'The economic theory of trade unions: An introductory survey', *The Scandinavian Journal of Economics*, vol. 87(2), pp. 160–193.

Picard, P. and Toulemonde, E. (2001). 'On the equivalence of taxes paid by employers and employees', *Scottish Journal of Political Economy*, vol. 48(4), pp. 461–470.

Pissarides, C.A. (2000). *Equilibrium unemployment theory*, MIT Press.

Riedl, A. and Tyran, J.R. (2005). 'Tax liability side equivalence in gift-exchange labor markets', *Journal of Public Economics*, vol. 89(11-12), pp. 2369–2382.

Robins, P.K. (1985). 'A comparison of the labor supply findings from the four negative income tax experiments', *Journal of Human Resources*, vol. 20(4), pp. 567–582.

Ruffle, B.J. (2005). 'Tax and subsidy incidence equivalence theories: experimental evidence from competitive markets', *Journal of Public Economics*, vol. 89(8), pp. 1519–1542.

Rupert, T.J. and Wright, A.M. (1998). 'The use of marginal tax rates in decision making: The impact of tax rate visibility', *Journal of the American Taxation Association*, vol. 20(2), pp. 83–99.

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*, vol. 127(1), pp. 493–533.

Sausgruber, R. and Tyran, J.R. (2005). 'Testing the mill hypothesis of fiscal illusion', *Public Choice*, vol. 122(1-2), pp. 39–68.

Sausgruber, R. and Tyran, J.R. (2011). 'Are we taxing ourselves?: How deliberation and experience shape voting on taxes', *Journal of Public Economics*, vol. 95(1-2), pp. 164–176.

Schokkaert, E. (2007). 'Capabilities and satisfaction with life', *Journal of Human Development*, vol. 8(3), pp. 415–430.

Schram, A. (2005). 'Artificiality: The tension between internal and external validity in economic experiments', *Journal of Economic Methodology*, vol. 12(2), pp. 225–237.

Schram, A.J. (1990). 'A dynamic model of voter behavior and the demand for public goods among social groups in Great Britain', *Journal of Public Economics*, vol. 41(2), pp. 147–182.

Shapiro, C. and Stiglitz, J.E. (1984). 'Equilibrium unemployment as a worker discipline device', *The American Economic Review*, vol. 74(3), pp. 433–444.

Simon, H.A. (1955). 'A behavioral model of rational choice', *The Quarterly Journal of Economics*, vol. 69(1), pp. 99–118.

Stiglitz, J. (2000). *Economics of the public sector*, Norton.

Sugden, R. (1984). 'Reciprocity: the supply of public goods through voluntary contributions', *The Economic Journal*, vol. 94, pp. 772–787.

Sutter, M. (2009). 'Individual behavior and group membership: Comment', *American Economic Review*, vol. 99(5), pp. 2247–2257.

Ullmann, R. and Watrin, C. (2008). 'Comparing direct and indirect taxation: The influence of framing on tax compliance', *European Journal of Comparative Economics*, vol. 5(1), pp. 23–56.

Weber, M. (2011). 'Income tax or employer's contribution – an experiment on the perception of 'economically equivalent' duties', *Tinbergen Institute Mphil-Thesis*.

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 2). 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 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:

◦

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 1 shows a screenshot of an employee's computer during the work task (taken in treatment *IN*). Figure 2 shows the subjective well-being self-assessment using the SAM-V-9 self-assessment manikin, and Figure 3 shows the slider that is used for the elicitation of preferences concerning the size of the public sector (taken in treatment *EN*).

C Appendix (for Online Publication): Sample Composition and Gender Effects

C.1 Sample Composition

Table C.1 shows the composition of the sample in the different treatments of the experiment.

Table C.1: Sample Composition

	EN	*IN*	*EP*	*IP*
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 4 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.

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 labour 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?'),

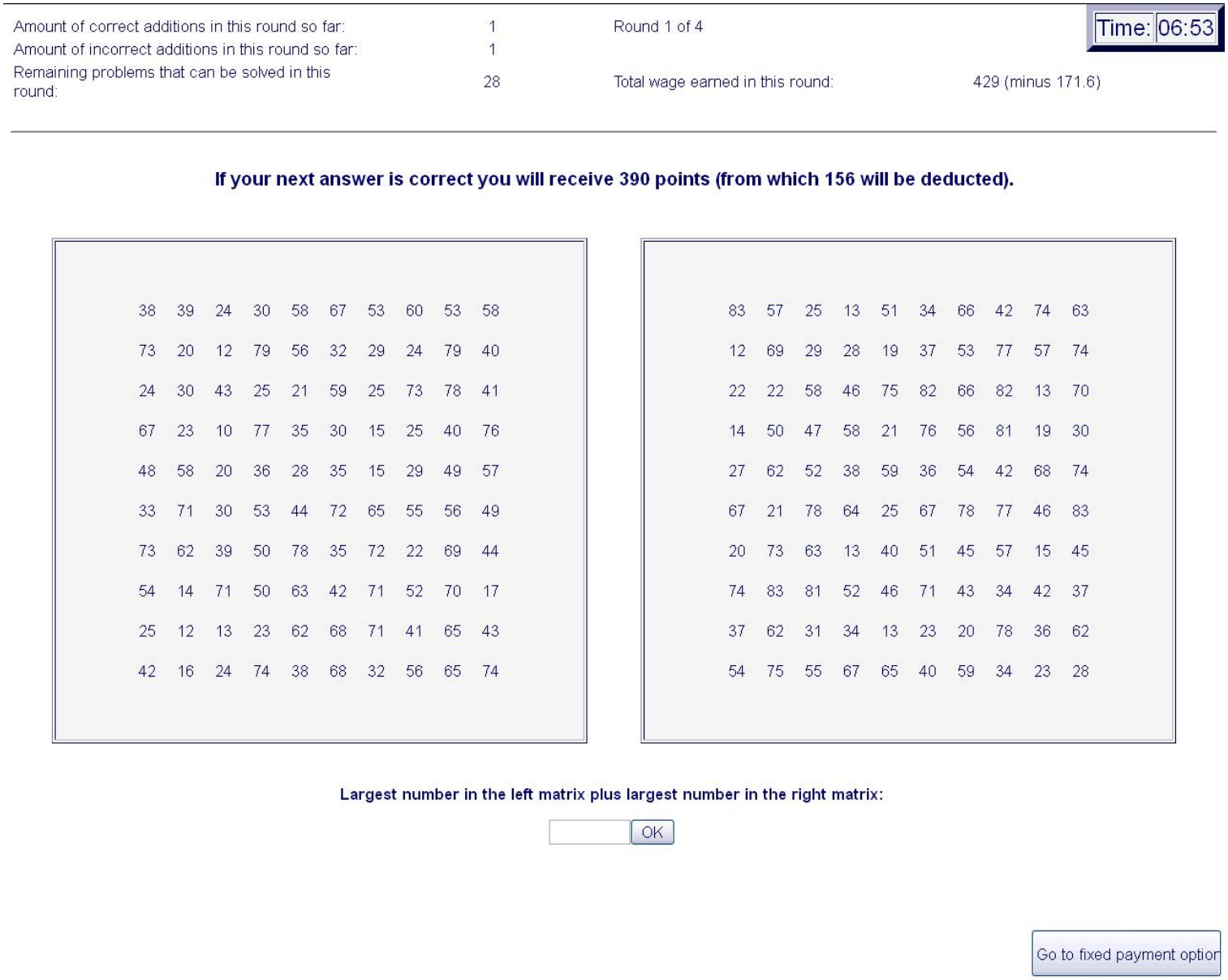

Figure 1: Screenshot during a Work Round

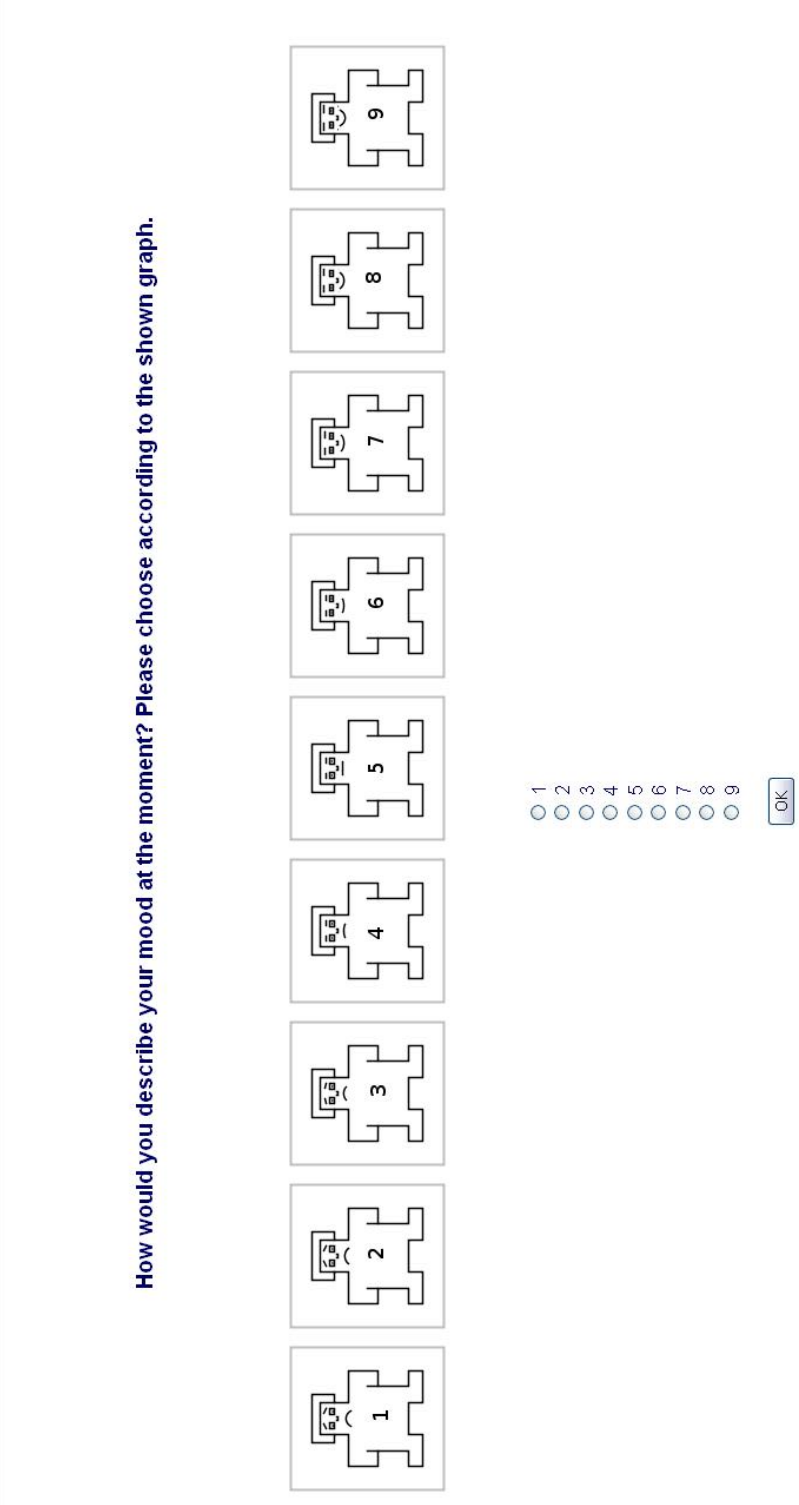

Figure 2: Screenshot during the Self-Assessment with the SAM-V-9

Figure 3: Screenshot of the Slider; Public Sector Size Preference

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 4 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 only use non-parametric tests here, separately for nothing in return and public good treatments. We split the data according to gender and compare the treatment differences among men and women.

Table C.2 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. Of 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*.[1]

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 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 labour 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 labour supply at the extensive margin (with and without public good, respectively) and marginally significant differences for labour supply at the intensive margin and job performance in the treatments with public good. When it comes down to such labour market responses, women are barely affected by the framing of the tax.

[1]The two observations that have missing data due to computer problems during the experiment are both from male participants (see Footnote 23 in the main text). 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.

Table C.2: Results Split According to Gender

	Employer payroll tax mean m/f	Income tax mean m/f	Treatment differences p-value Wilcoxon m/f
ps			
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
sw			
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
*le*1			
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
*le*2			
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
li			
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
jp			
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. The p-values stem from two-sided Wilcoxon rank-sum tests, comparing treatment effects separately for men and women. We denote by *ps* the preference concerning the size of the public sector; by *sw* the subjective well-being; by *le*1 and *le*2, respectively, the first and second measures of labour supply at the extensive margin; by *li* the labour supply at the intensive margin; and by *jp* the job performance.

While the results on public sector size preferences are somewhat inconclusive concerning gender effects, the results on subjective well-being and labour supply suggest that net wage illusion is stronger for men while the tax loss effect 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 the tax loss effect is the dominant mechanism. On the other hand, men are more sensitive to framing concerning the variables traditionally of most interest to economists – labour supply and job performance where net wage illusion is more important.

References (Online Appendix)

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74.

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.