

TI 2013-030/I  
Tinbergen Institute Discussion Paper



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

*Matthias Weber*  
*Arthur Schram*

*CREED, Faculty of Economics and Business, University of Amsterdam, and Tinbergen  
Institute.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 8579

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

Matthias Weber<sup>†</sup>

Arthur Schram<sup>†</sup>

February 18, 2013

## Abstract

In a competitive market with taxed transactions, it does not matter under full rationality which side of the market legally transfers the taxes. In the labor market, a tax levied on employers and a corresponding income tax levied on employees are equivalent. With boundedly rational agents, this equivalence is no longer obvious. If people react differently to the two taxes this has direct impact on policy making, political economics, and optimal taxation theory. This paper examines how people react to these duties in a real effort laboratory experiment. We study the differential effects of the two types of taxes on preferences concerning the size of the public sector, subjective well-being, labor supply, and on-the-job performance. To elicit public-sector-size preferences in the laboratory we introduce a novel, incentive compatible approach. Our findings suggest that employer-side taxes induce preferences for a larger public sector. Our findings also suggest that subjective well-being is higher while both labor supply and on-the-job performance are lower when the taxes are levied on employers. Furthermore, there are gender effects, e.g., women's subjective well-being appears to be more sensitive to framing than men's, while men's labor supply is more sensitive to framing than women's.

JEL classification: C91, H22, H30

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

---

\*Financial support by The Netherlands' Organisation for Scientific Research (NWO grant 406-11-022) and the Research Priority Area Behavioral Economics of the University of Amsterdam is gratefully acknowledged. Thanks for suggestions and comments go to Boris van Leeuwen, Pedro Robalo, Joep Sonnemans, Johannes Vatter, and participants of the ESA conference, New York, and the CREED-CEDEX-CBESS meeting, Norwich.

<sup>†</sup>CREED, University of Amsterdam and Tinbergen Institute

# 1 Introduction

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

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

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

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

---

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

<sup>2</sup>A small part of the traditional public finance literature allows for failure of LSE in the labor market due to market frictions. These studies maintain the assumption that economic agents react rationally to taxes. We are not interested here in studying labor market frictions, but in perceptions of and reactions to taxes that are not necessarily rational. Therefore, we study these taxes in a setting where they are by design equivalent under full-rationality, i.e. in a setting without frictions. Rejecting rational reaction

More specifically, we investigate whether people perceive distinct labor taxes differently that are equivalent under full rationality and whether they react to them in different ways. The labor market is an important market for studying behavioral responses to taxation, because deadweight losses from distortionary taxes on labor may be substantial (e.g. Feldstein, 1999). Two ways of taxing labor prevail around the world. One is an income tax levied on employees, the other a payroll tax levied on employers.<sup>3</sup> Note that these two types of taxes exist side by side in many countries (often in the form of contributions to social security on both the employee's and the employer's side). This is somewhat surprising from a full rationality point of view (at least, in the absence of market frictions), because under equivalence one would expect the tax to be chosen that minimizes collection and compliance costs. The reason for the co-existence is possibly that people perceive the two taxes differently and react to them in different ways. This is what our paper investigates.

There are many ways in which such taxes may differentially affect people. Here, we list three. First, there may be strong effects on individual political preferences. If perceptions vary across taxes of how much of the public sector is financed by distinct groups in society, opinions on the preferred size of the public sector (or welfare state) are likely to vary as well. This could directly influence voting decisions, which might explain why right-wing politicians tend to favor duties levied on the employees' side while left-wing politicians tend to favor duties levied on the employers' side.

Second, many economists are interested in the effects of policies on some index of well-being (representing individual utility or its aggregate, social welfare). Subjective well-being is therefore an obvious measure of the consequences of policies, including taxes.<sup>4</sup> People simply might be happier under some tax regimes than under others. It is possible, for example, that people get frustrated if they observe a large part of their gross wage being taxed away.

---

to taxes in such a setting also raises doubts about the underpinnings of models based on full rationality combined with market frictions.

<sup>3</sup>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, however. In the latter case, employees usually receive an entitlement that they do not receive with a tax. Nevertheless, on a theoretical basis, such duties can often be treated as taxes, because the same outcome can be achieved with taxes instead of contributions. We treat the terms as equivalent here. We use the term 'income tax' for a tax (or contribution) on the employees' (supply) side of the labor market and the term 'employer payroll tax' for a tax (or contribution) on the employers' (demand) side.

<sup>4</sup>For an overview of the literature on the measurement of subjective well-being, see Kahneman and Krueger (2006).

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

We consider the effects of differential perception of theoretically equivalent labor market taxes on each of these three dimensions: political-economic preferences, subjective well-being, and labor supply (together with performance). For this purpose, observational field data are ill-suited, because it is generally difficult to disentangle the numerous effects stemming from broad tax reforms. It is also often impossible to filter out the causes of observed effects (e.g., differences could stem from market frictions or from differences in perception) and moreover, counterfactuals are missing in such data. In addition, field experiments on taxes are almost impossible to implement as governments are highly unlikely to agree to implement a treatment design including proper controls, because not all citizens would be treated equally.<sup>5</sup>

This leaves laboratory experiments as a natural choice to investigate the questions at hand. Even if other empirical methods were feasible, for various reasons such experiments would still be a preferred way to investigate this issue. For one thing, the laboratory allows one to provide a setting that is most favorable for liability side equivalence to hold. All tasks, payoffs, and taxes are more salient and more directly related to decisions than is typically the case outside of the laboratory. Furthermore, institutional frictions are absent and laboratory control allows one to make the taxes equivalent by design instead of being equivalent only in general equilibrium.<sup>6</sup> 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

---

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

<sup>6</sup>In addition, the subject pool of many laboratory experiments, including ours, consists mainly of students, which are to a large part from the faculty of economics and business. They are on average more intelligent and better trained to understand taxation issues than the general public. This makes it more likely that the rationality assumption underlying LSE will be fulfilled.

opportunity to directly measure the effects of taxation. We will obtain individual-level measures of preferences for the size of the public sector, subjective well-being and labor supply responses.

In sum, we examine in a laboratory experiment with human subjects and monetary incentives whether people react differently to an incentive scheme depicting an income tax than to one reflecting a payroll tax levied on employers. By design, both duties are absolutely equivalent under full rationality.<sup>7</sup> To increase the external validity of our laboratory environment, the experiment will require real effort by subjects to earn an income (that may subsequently be taxed). As mentioned above, we will distinguish between two environments. In one, tax revenue does not benefit employers or employees in any way; in the other, proceeds are used to produce a public good that is supplied to all employees.<sup>8</sup> The public good is not supplied to employers because we envisage a public good related to income security, e.g., unemployment benefits. One could argue that the real world is an intermediate case – taxes yield something in return, but the returns perceived by any single tax payer are not always as obvious and direct as is the case of a public good supplied to a small group of subjects in the laboratory. Distinguishing between the two extreme cases allows us to isolate the effects on LSE of (perceived) returns from taxation.

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

This paper intends to add to the literature in the following ways.<sup>8</sup> 1. It is the first to investigate how levying a labor tax on either the employers' or the employees' side of the labor market influences individuals' preferences concerning the size of the public sector. 2. It is the first to investigate the effects of the liability side of a labor tax on subjective well-being. 3. It provides further evidence on the effects of the liability side of a labor tax on labor supply and job performance, being the first to do so in an environment that mimics the employer-employee relationship. 4. To elicit individuals'

---

<sup>7</sup>We will not mention this equivalence every time we compare the taxes. Throughout this paper, when we mention these taxes, we refer to a tax on the supply side and on the demand side of the labor market that are designed such that they are fully equivalent.

<sup>8</sup>Statements about this research being 'the first' are to the best of the authors' knowledge.

preferences concerning the size of the public sector in the laboratory, we introduce a novel, incentive compatible approach.

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

## 2 Related Literature

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

There are, however, other studies examining labor supply or job performance under such taxes. A part of this literature is theoretical.<sup>9</sup> 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. (2011) who investigate how gross earnings change when income tax rates or payroll tax rates change using recent French data. They find that gross labor earnings respond to changes in the marginal income tax rate while they do not respond to changes in the payroll tax rate, thus rejecting LSE. The authors suggest that this might be due to differential effects of these tax changes on labor supply. Using data from the Netherlands Muysken et al. (1999) present evidence that a larger part of taxes is shifted if they are levied on the employees' side rather than on the employers' side, Holm et al. (1994) use Finnish data in an empirical application of a monopoly union wage determination model and find that increasing the payroll tax

---

<sup>9</sup>There are different approaches to modeling the labor market and thus the impact of labor market taxes. Most prominent are the competitive labor market approach (see e.g. Atkinson and Stiglitz, 1980), the efficiency wage theory (see Shapiro and Stiglitz, 1984), search and matching models (see Pissarides, 2000) and union bargaining models (see e.g. Oswald, 1985). Most of this literature does not allow for liability side non-equivalence. Such non-equivalence can arise in exceptional cases via market frictions, however (see e.g. Koskela and Schöb, 1999, Picard and Toulemonde, 2001, Rasmussen, 1997, Rasmussen, 1998). In general, none of these approaches allow for any non-equivalence due to tax perception. In the presence of labor market frictions, it seems likely that non-rational perception of taxes can considerably amplify or dampen the effects of these frictions.



rate has a negative effect while increasing the income tax rate has a positive effect on wages.

There is a limited experimental literature on the effects of tax liability on labor supply. We know of no such study implementing an employer-employee relationship in the laboratory. In a setting without employers, Gamage et al. (2010) experimentally study distinct income tax schemes (specifically, no tax, a flat tax, a progressive tax and a wage subsidy) while adjusting the gross wages such, that all treatments are equivalent under full rationality. Their results show that labor supply differs across treatments. Their most robust finding is that subjects choose to supply less labor in the wage subsidy treatment than in the others. Fochmann et al. (2010b) investigate whether the gross wage has an influence on the labor supply decisions of the participants. They find that participants choose to work longer and harder when their gross wage is higher (holding net wages constant). Their experiment distinguishes between a no tax treatment, a 25% tax treatment and a 50% tax treatment. They find lower labor supply in the no tax treatment than in the two tax treatments and refer to their finding as “gross wage illusion”.

In addition to the literature on tax liability and labor supply, there is some experimental research investigating labor supply responses to other ‘equivalent’ taxes. Blumkin et al. (2012) examine labor supply reactions to an income tax and to a consumption tax that are equivalent under full rationality. They find that experimental subjects faced with a proportional income tax work significantly less than those faced with a corresponding consumption tax. Fochmann and Weimann (2011) elaborate on Fochmann et al. (2010b) and explain gross wage illusion by tax salience. In their model the misperception of taxes depends on the true tax rate – they assume that the tax bias is zero for tax rates of 0% and 100%, while tax rates can be misperceived in between. They add an additional experiment showing that tax salience matters for effort provision under different progressive income tax schemes.

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. Kerschbamer and Kirch-

steiger (2000) investigate liability side equivalence in an ultimatum game setting, and find that it does not hold. Riedl and Tyran (2005) investigate gift exchange markets. Their results support LSE. Finally, Cox et al. (2012) examine tax incidence in double auction and posted offer markets and find that LSE does not hold.

All in all, the experimental work on LSE in various environments provides mixed results, though most studies report violations. Considering tax perception more generally, many studies report seemingly irrational behavior by laboratory participants. An excellent survey is presented in Fochmann et al. (2010a). An example of this literature is De Bartolome (1995), who shows that many people mistakenly use the average tax rate instead of the marginal tax rate when making investment decisions. Fochmann et al. (2012) study how investment decisions change with the framing of taxes. Their experimental results show that the possibility to deduct losses from an income tax leads to significantly riskier investments (again, their treatments are equivalent under full rationality). 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.<sup>10</sup>

### 3 Experimental Design and Procedures

The experiment was conducted at the CREED laboratory at the University of Amsterdam in February 2012 with a total of 240 subjects recruited from the CREED subject pool.<sup>11</sup> The experiment was programmed in PHP/MySQL. Every participant received a

---

<sup>10</sup>There are other studies that are more loosely related to our research. These include Blaufus et al. (2010), who conduct a survey to investigate whether subjects rationally perceive the interaction of tax rate and tax base (finding that this is not the case). These authors also simulate voting outcomes for different tax rate/tax base combinations, finding that subjects would not choose the rational combinations. McCaffery and Baron (2006) give an overview of different heuristics and biases that are present when people are faced with taxes. McCaffery (1994) elaborates on different behavioral aspects of tax perception (and gives nice examples of how some of these are already considered in existing tax laws). Reimers (2009) finds that if individuals are asked to determine a fair tax, they choose more progressive taxes when the tax is expressed in percentage terms rather than in absolute terms. The same holds when the tax it is expressed as post-tax money retained rather than as tax paid. Esarey et al. (2012a) and Esarey et al. (2012b) examine in different settings how political preferences concerning redistribution arise in the laboratory. O’Donoghue and Rabin (2003) examine how optimal taxation theory is affected when people are not fully rational, but prone to self-control problems.

<sup>11</sup>Participants were mainly undergraduate students. Slightly less than half were female and roughly 60% were majoring in economics or business.

show up fee of 7 euros. During the experiment, ‘points’ were used as currency. These were exchanged into euros at the end of each session at an exchange rate of 1 euro per 600 points. The experiment lasted between 90 minutes and 2 hours in total and participants earned on average about 22 euros, including the show up fee. Before starting, the participants had to answer control questions to make sure that they understood the instructions. The experiment did not start until all participants had successfully answered these questions. 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.<sup>12</sup> When scheduling, we distributed the sessions in a balanced way over mornings and afternoons and over the different days of the week. Appendix A provides a transcript of the instructions and the test questions, Appendix B contains screenshots.

### 3.1 Treatments

The design is a  $2 \times 2$  factorial, between-subject design. Subjects are either employer or employee. Those in the role of employees work on a task for which they receive performance-based remuneration. The form this incentive scheme takes is one of the treatment variables. In one case, employees receive a gross wage, from which a duty is subsequently deducted as a tax. In the other, employers pay the duty and employees receive a (lower) net wage. Note that this corresponds naturally to labor market taxes levied either on the employees’ or on the employers’ side. What happens with the tax proceeds is varied in the second treatment variable. The tax proceeds are either taken away (‘nothing in return’) or used to produce a public good (which is called ‘common fund’ in the experiment). Table 1 summarizes the design. The acronyms *EN*, *IN*, *EP* and *IP* for the four distinct treatments will be used regularly in the remainder. In parentheses are the numbers of subjects per treatment.<sup>13</sup>

The tax rate used is 40% in the income tax treatments and 66.7% of the corresponding

---

<sup>12</sup>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.

<sup>13</sup>A quick rule of thumb power calculation yields that the number of observations per treatment is large enough to detect a 0.6 standard deviation treatment effect with probability 80% (at significance level 5%, assuming normality and two sided tests; using one-sided tests yields of course higher power while using non-parametric tests yields lower power; we use almost exclusively data from employees, which constitute  $\frac{5}{6}$  of all observations).

Table 1: 2×2 design

|                   | employer payroll tax | income tax     |
|-------------------|----------------------|----------------|
| nothing in return | <i>EN</i> (60)       | <i>IN</i> (60) |
| public good       | <i>EP</i> (60)       | <i>IP</i> (60) |

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

lower wage in the employer payroll tax treatments. This ensures equivalence between the income tax and the employer payroll tax in terms of employees' net earnings and employers' gross costs, both for the nothing in return and the public good treatments. Hence, differences between *EN* and *IN* and between *EP* and *IP* reflect only framing effects. The wages were also chosen such that the net wage in a nothing-in-return treatment is equal to the net wage plus the return from the public good from one's own tax payment in the public good treatment.<sup>14</sup> Note that this means that average total net earnings are higher in the public good treatments than in the treatments without public good, because employees also benefit from returns of the public good that are due to taxes paid by other employees in their group.

The public good in the relevant treatments is produced with a multiplication factor of 1.3 (meaning that tax revenue increased by 30% is allocated to the common fund) and its returns are equally distributed among all employees in a group at the end of the experiment.<sup>15</sup>

<sup>14</sup>Recall that in the public good treatments the net earnings consist of the net wage plus the returns from the public good. The returns from the public good can be split in a part that is due to own tax payment and a part that is due to the taxes paid by the other employees in the group. Consider a task an employee faces. Denote by  $W_P$  and  $W_N$  the net wages for this task in the public good and the nothing in return treatments, and by  $\gamma \cdot W_P$  the tax paid, which is a (mandatory) contribution to the public good, where applicable. In the nothing in return treatments the employee receives  $W_N$ , which compares to  $W_P + \frac{\gamma \cdot W_P \cdot 1.3}{5}$  in the public good treatments.  $W_P$  and  $W_N$  are chosen to equate these two returns. Finally, note that subjects receive no information on the choices or performances of others during the experiment. Hence, their choices cannot influence others.

<sup>15</sup>As noted above, employers do not benefit from the public good. As a consequence, a tax revenue of  $x$  points in a group with  $n$  employees, yields  $\frac{1.3x}{n}$  points from the public good for each employee.

## 3.2 Course of Events

Figure 4 at the end of this section summarizes the design by giving a schematic overview of the timeline. In 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, terms such as ‘income tax’ or ‘employer payroll tax’ are not used to avoid (unmeasured) preconceptions that some subjects might have with respect to these terms.

The experiment involves a real-effort work task which is the following. Each employee sees two  $10 \times 10$  matrices on the computer monitor that are filled with randomly generated two-digit numbers. Figure 1 shows a screenshot from the work task (taken from treatment *IN*).<sup>16</sup> One matrix is shown on the left half of the screen and the other one on the right half. The employees’ job is to find the largest number in the left matrix and the largest number in the right matrix and to add these two numbers up. For the summation, the participants are provided with pocket calculators. After answering, irrespective of whether the answer is correct, a new pair of matrices appears. This means that subjects have only one attempt to provide the correct answer. Each employee faces a maximum of 30 of these problems, which is much more than they can actually solve correctly. This limit and the way the random numbers are generated make guessing a very unsuccessful strategy. Only the number of correct additions matters, there is no punishment for incorrect additions. While the employees are doing this task, they can see at the top of the monitor the amount they will receive if the next number they enter is correct and, where applicable (i.e., when tax is levied on the employees), the amount that will be deducted from it (as a tax). Furthermore, they can see how much they have already earned and, where applicable, the amount that will be deducted from it. They can also see the number of correct and incorrect additions so far and the remaining time. This procedure is repeated in four independent and identical rounds, each lasting for 8 minutes. During these rounds, employers do not need to do anything. Note that the total number of correctly solved problems is a measure of job performance.

Employers receive a net payment of 49.8 points per correct addition by any of their five

---

<sup>16</sup>A larger version of this screenshot can be found in Appendix B.

Figure 1: Screenshot during a work round

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

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

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 38 | 39 | 24 | 30 | 58 | 67 | 53 | 60 | 53 | 58 |
| 73 | 20 | 12 | 79 | 56 | 32 | 29 | 24 | 79 | 40 |
| 24 | 30 | 43 | 25 | 21 | 59 | 25 | 73 | 78 | 41 |
| 67 | 23 | 10 | 77 | 35 | 30 | 15 | 25 | 40 | 76 |
| 48 | 58 | 20 | 36 | 28 | 35 | 15 | 29 | 49 | 57 |
| 33 | 71 | 30 | 53 | 44 | 72 | 65 | 55 | 56 | 49 |
| 73 | 62 | 39 | 50 | 78 | 35 | 72 | 22 | 69 | 44 |
| 54 | 14 | 71 | 50 | 63 | 42 | 71 | 52 | 70 | 17 |
| 25 | 12 | 13 | 23 | 62 | 68 | 71 | 41 | 65 | 43 |
| 42 | 16 | 24 | 74 | 38 | 68 | 32 | 56 | 65 | 74 |

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 83 | 57 | 25 | 13 | 51 | 34 | 66 | 42 | 74 | 63 |
| 12 | 69 | 29 | 28 | 19 | 37 | 53 | 77 | 57 | 74 |
| 22 | 22 | 58 | 46 | 75 | 82 | 66 | 82 | 13 | 70 |
| 14 | 50 | 47 | 58 | 21 | 76 | 56 | 81 | 19 | 30 |
| 27 | 62 | 52 | 38 | 59 | 36 | 54 | 42 | 68 | 74 |
| 67 | 21 | 78 | 64 | 25 | 67 | 78 | 77 | 46 | 83 |
| 20 | 73 | 63 | 13 | 40 | 51 | 45 | 57 | 15 | 45 |
| 74 | 83 | 81 | 52 | 46 | 71 | 43 | 34 | 42 | 37 |
| 37 | 62 | 31 | 34 | 13 | 23 | 20 | 78 | 36 | 62 |
| 54 | 75 | 55 | 67 | 65 | 40 | 59 | 34 | 23 | 28 |

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

Go to fixed payment option

employee's. Net payments for the employees are linearly decreasing in the number of attempts (but are restricted to be non-negative). If employees solve the first problem correctly in the nothing in return treatments, they receive 280.8 points. With each attempt (whether correct or not) the payment for the next correct addition decreases by 23.4 points.<sup>17</sup> In the public good treatments, these numbers correspond to the net return from own performance (the direct net wage plus the return from the public good that is due to own performance, for the calculation see Footnote 14).

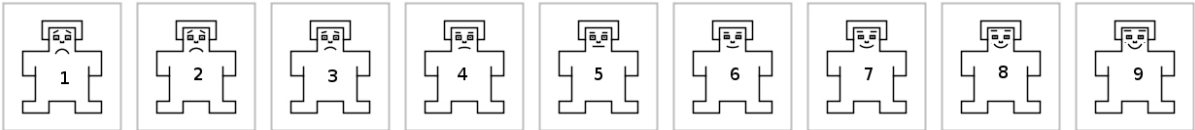
We provide employees with an outside option. Instead of working, they can also choose

<sup>17</sup>Formally, net earnings from correctly solving a problem in the nothing in return treatments can be written as  $\pi = \max(280.8 - 23.4x, 0)$ , where  $x$  is the number of problems the employee has previously attempted to solve in the same round. This is also the gross wage in the treatment *EN* (net wage equals gross wage), while the gross wage in the treatment *IN* is  $\pi = \max(468 - 39x, 0)$ , which leads, with a tax rate of 40%, to the same net wage as in *EN*.

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

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. Next, participants are surveyed to measure their subjective well-being using a self-assessment manikin (SAM), the SAM-V-9 (Irtel, 2007).<sup>18</sup> 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 2.

Figure 2: 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.

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.<sup>19</sup> 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).<sup>20</sup> It is randomly determined whether the extra round takes place or not. If it takes place and if the price is lower than the amount stated by an employee, this employee pays the price and works

<sup>18</sup>See Lang (1985), or Bradley and Lang (1994). This subjective well-being is sometimes referred to as satisfaction, happiness, or experienced pleasure.

<sup>19</sup>Because this is asked before some are selected to be employers, all provide a willingness to pay; for those randomized to be employers the revealed willingness to pay has no further consequence.

<sup>20</sup>The lower limit 1056 corresponds to the amount earned after immediately choosing the fixed payment option). The upper limit 2400 is a number slightly higher than the expected maximum possible earnings in one round.

(and gets paid) for another round. If the price is higher than the bid of the employee, the employee neither pays for nor works in another round. Subjects not participating remain seated until all participants have finished. If subjects have a true valuation for participating in this extra round it is a dominant strategy to bid this true valuation. In the treatments with public good the returns from the public good are split among all employees of a group, those working in the extra round and those not working. Note that the stated willingness to pay provides us with a (first) measure of labor supply at the extensive margin.

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

After this second statement of willingness to pay, subjects are told that there will be yet another round. All employees participate in this round, which will take place with new rules. The rules differ from the regular rounds as follows. In the nothing in return treatments (*EN* and *IN*), a public good is introduced, such that the taxes are no longer lost, just as in the public good treatments. Now, the tax rate and the multiplication factor for the public good are no longer given. Instead, they are chosen by the employees in a random dictator style, using the following mechanism. Subjects are presented with a slider, as shown in Figure 3. 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.<sup>21</sup> 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

---


<sup>21</sup>The trade-off between tax rate and multiplication factor can be interpreted as a diminishing marginal productivity of the public sector; the higher the tax revenue is, the lower is the efficiency of public good production. While one could also have subjects choose the tax rate for a fixed multiplication factor, we prefer to explicitly allow for this trade-off in order to accommodate preferences for a point at the interior of the slider (with fixed multiplication factor, one would expect many corner solutions where subjects prefer either 0 or maximal tax rates).



Figure 3: Slider for the elicitation of public sector size preference

**Please choose your most preferred position of the slider.**

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



OK

*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.

default position.<sup>22</sup> The number corresponding to the chosen slider position provides us with a measure of the subject’s preferred size of the public sector. After all subjects have chosen a slider position, one employee in each group is randomly selected and her choice is used for this extra round.<sup>23</sup>

After having chosen a position of the slider, subjects are offered different lotteries in order to elicit information on their loss aversion in risky choices. We use the same lotteries as Fehr and Goette (2007). For each of two lotteries, subjects can decide whether or not they want to participate. In lottery *A*, participants can either win 4 euros or lose 2.5 euros with equal probabilities. Lottery *B* consists of six independent repetitions of lottery *A*. One of the lotteries *A* and *B* is chosen randomly and carried out if the subject decided to participate in it.<sup>24</sup>

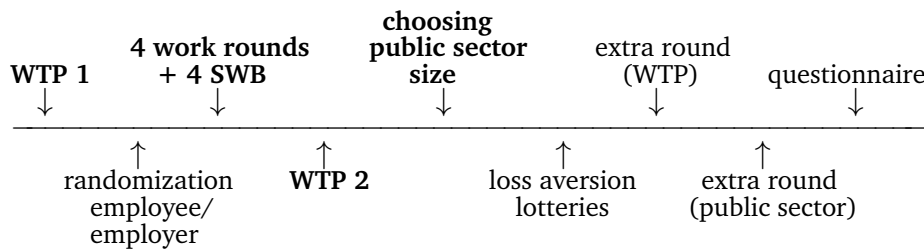
<sup>22</sup>The multiplication factor of the public good is 3 at the default position on the left end and 0.75 at the right end. The tax is 0 at the left end and 75% of the employer’s labor cost at the right end; the tax is always expressed as a percentage of the employee’s (gross) wage.

<sup>23</sup>Note that the employer payroll tax and the income tax treatments are still absolutely equivalent, whereas the nothing and return and the public good treatments are now somewhat ‘less equivalent’. Subjects in the nothing in return treatments have had no experience with the public good prior to this round. Furthermore, net payments in the nothing in return treatment are not adjusted to the levels used in the public good treatment in order to avoid subjects having to adapt to a new payment schedule for the extra round. As a consequence, payoffs here are slightly higher in the treatments *EN* and *IN* than in *EP* and *IP*.

<sup>24</sup>Subjects who reject both lotteries are more loss averse than subjects who reject only lottery *A*. Subjects who reject lottery *A* are more loss averse than subjects who accept both lotteries. Accepting lottery

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). The information on the randomly drawn price and the public sector size parameters selected in random dictator style is given to subjects before the respective round starts. After these two rounds information on payments stemming from others' public goods contributions (i.e., their tax payments) and on the outcomes of the lotteries are revealed. Before being paid, participants are asked to fill out a short questionnaire, including questions concerning gender, age, field of study and experience in lab experiments.<sup>25</sup>

Figure 4: Timeline of the experiment



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

## 4 Hypotheses

The theory underlying this experimental paper is straightforward. The null-hypotheses for all dependent variables we consider arise from the behavior of fully rational economic agents. The incentive schemes we use are equivalent by design, which means that no calculations or general equilibrium mechanisms are necessary to arrive at the

A while rejecting lottery B is inconsistent. These claims hold if individuals maximize reference dependent utility as in Goette et al. (2004) and Fehr and Goette (2007) that simplifies to maximizing  $v(x-r) = \begin{cases} \lambda(x-r) & \text{if } x \geq r \\ \gamma\lambda(x-r) & \text{if } x < r \end{cases}$  where  $x$  is the outcome of the gamble,  $r$  is the reference point (zero in our case), and  $\gamma > 1$  measures loss aversion ( $\lambda > 0$ ). The proof of these claims can be found in Fehr and Goette (2007).

<sup>25</sup>The questionnaire is reproduced in Appendix A.

full-rationality outcome. All null-hypotheses reflect no difference in outcomes between the respective employer payroll tax and income tax treatments.

Consider the dependent variables derived from subjects' choices as described in Section 3.2. We define variables straightforwardly, denoting by  $ps$  the preference concerning the size of the public sector (as measured by the chosen slider position); by  $sw$  the subjective well-being (as measured by the four choices with the self-assessment manikin); by  $le1$  and  $le2$ , respectively, the first and second measures of labor supply at the extensive margin (as given by choices in the two BDM tasks); by  $li$  the labor supply at the intensive margin (measured by the number of seconds in the work mode over the four regular rounds); and by  $jp$  the job performance (measured by total output). We denote null-hypotheses by  $H_{.,0}$  and alternative hypotheses by  $H_{.,A}$ .

Formulating alternatives to the nulls, we start with preferences regarding the size of the public sector. An intuitive alternative to full rationality is that people favor a larger public sector when they feel that someone else is paying for it. Subsequently, we hypothesize that employees are more likely to sense that someone else is paying when taxes are levied on the employers' side of the labor market. Thus, the alternative to full rationality is that employees prefer a larger public sector in  $EN$  than in  $IN$  and in  $EP$  than in  $IP$ . We thus test:

$$H_{1,0} : ps_{EN} = ps_{IN} \text{ vs } H_{1,A} : ps_{EN} > ps_{IN} \text{ and} \\ H_{2,0} : ps_{EP} = ps_{IP} \text{ vs } H_{2,A} : ps_{EP} > ps_{IP}.$$

Next, consider subjective well-being. In the nothing in return treatments, the most intuitive alternative is that people are less happy when they feel that the tax represents money taken directly from them. This leads to the test:

$$H_{3,0} : sw_{EN} = sw_{IN} \text{ vs } H_{3,A} : sw_{EN} > sw_{IN}.$$

In the public good treatments one may expect a similar effect. In case of subjects with other-regarding preferences, the effect may be weaker, however. This is because the 'pain' of having something taken away may be mitigated by a feeling of doing something good for others (by contributing to the public good). In extreme cases this could even reverse the effect. Hence, we have no directional alternative hypotheses and test:

$$H_{4,0} : sw_{EP} = sw_{IP} \text{ vs } H_{4,A} : sw_{EP} \neq sw_{IP}.$$

Finally, for labor supply and job performance, various outcomes are conceivable. A

likely alternative to full rationality is what Fochmann et al. (2010b) call ‘gross wage illusion’ (as observed by Fochmann et al. (2010b) and Gamage et al. (2010) in different settings). This implies that people supply more labor in the income tax treatments, because they focus on before-tax income when making their decisions. This alternative hypothesis is consistent with reference dependent utility maximization of loss-averse individuals if the reference earning in a round is determined by the gross earning arising from the performance individuals expect of themselves.<sup>26</sup> For subjects with other-regarding preferences, the feeling of helping others (by contributing to the public good) may amplify the effect in the treatments with a public good. On the other hand, one can also conceive opposite effects. For example, subjects may dislike the explicit deductions from income that characterize the income tax treatments and therefore decide to work less. We refrain from attributing weights to these alternative hypotheses and thus test undirected alternative hypotheses. We test

$$\begin{aligned}
H_{5,0} : le_{1EN} = le_{1IN} \text{ vs } H_{5,A} : le_{1EN} \neq le_{1IN}, \\
H_{6,0} : le_{1EP} = le_{1IP} \text{ vs } H_{6,A} : le_{1EP} \neq le_{1IP}, \\
H_{7,0} : le_{2EN} = le_{2IN} \text{ vs } H_{7,A} : le_{2EN} \neq le_{2IN}, \\
H_{8,0} : le_{2EP} = le_{2IP} \text{ vs } H_{8,A} : le_{2EP} \neq le_{2IP}, \\
H_{9,0} : li_{EN} = li_{IN} \text{ vs } H_{9,A} : li_{EN} \neq li_{IN}, \\
H_{10,0} : li_{EP} = li_{IP} \text{ vs } H_{10,A} : li_{EP} \neq li_{IP}, \\
H_{11,0} : jp_{EN} = jp_{IN} \text{ vs } H_{11,A} : jp_{EN} \neq jp_{IN}, \text{ and} \\
H_{12,0} : jp_{EP} = jp_{IP} \text{ vs } H_{12,A} : jp_{EP} \neq jp_{IP}.
\end{aligned}$$

## 5 Results

We have collected data from 10 groups per treatment. Because we only use data obtained from employees, this gives 50 observations per treatment, except for the first measure of labor supply at the extensive margin, where we have 60 observations per treatment (as this was measured before the randomization into employees and employers).<sup>27</sup> Because subjects receive no feedback about others’ decisions until the end

<sup>26</sup>Behavior consistent with reference dependent utility maximization and loss aversion around a daily reference income has been found e.g. in Fehr and Goette (2007) and Crawford and Meng (2011); that expectations can influence reference points has been shown by (Abeler et al., 2011)

<sup>27</sup>Due to computer problems, one observation in *IN* has missing data on *sw* and one observation in *IP* has missing data on *sw*, *li*, and *o*.

of the experiment, we treat observations across individuals as statistically independent. We present here only the results of Wilcoxon rank-sum tests as these draw upon the least restrictive assumptions concerning the distribution underlying the data.<sup>28</sup> Tests performed for one-directional alternative hypotheses (specifically,  $H_{1,A}$ ,  $H_{2,A}$ , and  $H_{3,A}$ ) are one-sided; all other tests are two-sided.<sup>29</sup>

## 5.1 Public Sector Size Preference

Table 2 shows the means of the variable concerning the preference for public sector size ( $ps$ ) with standard errors and the p-values of Wilcoxon rank-sum tests. It shows that subjects in the employer payroll tax treatments prefer higher taxes than those in the income tax treatments. This is independent of the question whether subjects had previously experienced rounds without or with a public good: the differences are significant at the 1% and at the 5% level, respectively. Thus, we reject  $H_{1,0}$  in favor of alternative hypothesis  $H_{1,A}$  and  $H_{2,0}$  in favor of  $H_{2,A}$ . This supports the intuition described above that people prefer a larger public sector when they feel (irrationally) that someone else is paying for it.

Table 2: Results for public sector size preference

|  | employer<br>payroll tax<br>mean (std. error) | income<br>tax<br>mean (std. error) | treatment<br>differences<br>p-value Wilcoxon |
|--|--|------------------------------------|--|
| <b>public sector<br/>size preference</b> |  |                                    |  |
| nothing in return                        | 51.18 (4.22)                                 | 33.00 (2.77)                       | 0.001  |
| public good                              | 41.18 (3.69)                                 | 33.12 (2.96)                       | 0.040  |

*Notes:* The table shows the results for individuals' preferences for the size of the public sector (the variable  $ps$ ). Individual outcomes are integers between 0 and 100, larger numbers representing preference for a larger public sector.

<sup>28</sup>Our conclusions do not change when employing other tests or controlling for gender, age, field of study, experience with laboratory experiments or loss aversion. Details are available upon request.

<sup>29</sup>The data analysis was carried out in R (R Core Team, 2012).

## 5.2 Subjective Well-Being

Table 3 shows summary statistics and p-values of Wilcoxon rank-sum tests concerning subjects' subjective well-being ( $sw$ ). Our results show higher subjective well-being in the employer payroll tax treatment when tax proceeds are lost. We reject the null-hypothesis  $H_{3,0}$  that there is no difference at the 1% level. The results show no significant difference between the public good treatments. Hence, we cannot reject the null  $H_{4,0}$  in favor of the alternative  $H_{4,A}$ . Apparently, people feel unhappier about taxes, when they are levied on their gross income, but only if this money is lost. <sup>30</sup>

Table 3: Results for subjective well-being

|                                  | employer<br>payroll tax<br>mean (std. error) | income<br>tax<br>mean (std. error) | treatment<br>differences<br>p-value Wilcoxon |
|----------------------------------|--|------------------------------------|--|
| <b>subjective<br/>well-being</b> |  |                                    |  |
| nothing in return                | 21.84 (0.66)                                 | 18.94 (0.84)                       | 0.003  |
| public good                      | 21.16 (0.77)                                 | 21.82 (0.75)                       | 0.594  |

*Notes:* The table shows the results of individuals' subjective well-being (the variable  $sw$ ). Individual outcomes are integers between 4 and 36, larger numbers representing higher subjective well-being.

## 5.3 Labor Supply and Job Performance

Table 4 shows summary statistics and p-values for the labor supply variables and job performance (output). Our results yield for most variables (six out of eight cases) higher outcomes when the employee pays the tax, with or without a public good. Only two of these findings are statistically significant, however. First, when proceeds are used to produce a public good, labor supply at the extensive margin – measured using the BDM procedure after subjects had completed four rounds of the task – is significantly higher at the 5%-level when the tax is levied at the employee's side than when it is levied at the employer's side. Second, with public good, labor supply at the intensive margin is significantly higher at the 10%-level when the tax is levied at the employee's side. Because many of the other p-values are relatively low and the results for labor supply

<sup>30</sup>Recall that the our measurement of subjective well-being is the sum of four responses to the manikin scale we used. Consideration of the four separate self-reports shows very consistent behavior and no indication of trends. More information is available upon request.

and performance are higher with an income tax for these low p-values, we interpret the results as giving mild support to the idea that labor supply and output are higher when the tax is levied on the employee's side. Formally, we only reject the null hypotheses  $H_{8,0}$  and  $H_{10,0}$  in favor of  $H_{8,A}$  and  $H_{10,A}$ , respectively. The other null hypotheses on labor supply and job performance cannot be rejected. A noticeable effect in Table 4 is that labor supply at the extensive margin decreases substantially after subjects have experienced the task involved. The willingness to pay for an additional round decreases from the 1400-1500 range to the 1200-1300 range. We think that this might be due to subjects misjudging the difficulty of the task and the fatigue arising from doing this task for four rounds.

Table 4: Results for labor supply and job performance

|  | employer payroll tax<br>mean (std. error) | income tax<br>mean (std. error) | treatment differences<br>p-value Wilcoxon |
|--|---|---------------------------------|---|
| <b>labor supply extensive margin 1</b> |   |                                 |   |
| 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                                     |
| <b>labor supply extensive margin 2</b> |   |                                 |   |
| 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                                     |
| <b>labor supply intensive margin</b>   |   |                                 |   |
| 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                                     |
| <b>job performance</b>                 |   |                                 |   |
| nothing in return                      | 20.56 (1.07)                              | 19.64 (1.17)                    | 0.679                                     |
| public good                            | 18.40 (1.21)                              | 21.02 (1.23)                    | 0.139                                     |

*Notes:* The table shows the results of individuals' labor supply and performance (the variables  $le1$ ,  $le2$ ,  $li$ , and  $jp$ ). Higher numbers represent higher labor supply or performance. The p-values stem from two-sided Wilcoxon rank-sum tests.

## 5.4 An Explorative Analysis of Gender Effects

In the world outside of the laboratory, large gender differences exist regarding many economic variables, particularly labor supply (see Evers et al., 2008a). 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. Table 5.4 shows the results from the previous subsections, now split up according to gender. It is striking that some of the effects reported above are driven solely by male participants while some others are driven solely by female participants. Note that we are not interested in differences in levels here (e.g., do women produce more output?), but in differences in the treatment effects (e.g., is female job performance affected differently by tax framing than male job performance?). One thing to keep in mind with this analysis is that, of course, the sample sizes are now smaller. There are 32 (18) men (women) in treatment *EN*, 25 (25) in *IN*, 24 (26) in *EP*, and 31 (19) in *IP*.<sup>31</sup>

A first thing to observe is that the aggregate results concerning public sector size preferences are mainly driven by men when there is no public good while they are solely 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, but smaller, effect for women is only marginally significant. 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. Apparently women (irrationally) feel a higher burden when paying the taxes to fund the public good than men do.

In terms of subjective well-being, the aggregate result that when proceeds are wasted, people feel better if the tax was paid by the employer, is driven by women. The well-being reported by men barely responds to the framing of the taxes.

The labor supply and job performance effects are all mainly driven by men, with highly significant differences (p-values of 0.007 and 0.02) for the second measure of labor supply at the extensive margin and marginally significant differences for labor supply at the intensive margin and job performance in the treatments with public good. When it comes down to such labor market responses, women are barely affected by the framing of the tax. In contrast, men respond strongly to gross wages, even if they know that some of it will be taxed away. Gross wage illusion appears to make men work more and work harder.

Most of the economic literature looking at gender effects seems to suggest that women

---

<sup>31</sup>The two observations that have missing data due to computer problems during the experiment are both from male participants (see Footnote 27).



Table 5: Results, split up according to gender

|  | employer<br>payroll tax<br>mean m/f | income<br>tax<br>mean m/f | treatment<br>differences<br>p-value Wilcoxon m/f |
|--|-------------------------------------|---------------------------|--|
| <b>public sector<br/>size preference</b>   |                                     |                           |  |
| nothing in return                          | 54.44/45.39                         | 31.84/34.16               | 0.004/0.073                                      |
| public good                                | 40.46/41.85                         | 36.68/27.32               | 0.317/0.012                                      |
| <b>subjective<br/>well-being</b>           |                                     |                           |  |
| nothing in return                          | 21.66/22.17                         | 20.71/17.24               | 0.364/0.002                                      |
| public good                                | 22.42/20.00                         | 21.67/22.05               | 0.814/0.351                                      |
| <b>labor supply<br/>extensive margin 1</b> |                                     |                           |  |
| 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                                      |
| <b>labor supply<br/>extensive margin 2</b> |                                     |                           |  |
| 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                                      |
| <b>labor supply<br/>intensive margin</b>   |                                     |                           |  |
| 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                                      |
| <b>job performance</b>                     |                                     |                           |  |
| 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 results split by gender. Larger numbers represent preference for a larger public sector, higher subjective well-being, supply of more labor, and higher performance, respectively.

are more sensitive to framing than men (see Croson and Gneezy, 2009). In contrast, our results suggest that women are more sensitive to framing for some measures (in particular, subjective well-being) while men are more sensitive to framing for others. Concerning the variables traditionally of most interest to economists – labor supply and job performance –, we find men to be more sensitive to framing than women.

## 6 Policy Implications

The policy implications of our results are non-trivial. These show that who formally pays the tax affects individuals in at least three ways: their political preferences, their subjective well-being and their labor supply (including output). Under LSE, none of these should be affected, because it does not matter who pays. Each of these effects may have consequences for social welfare and/or policy implications. Simultaneously considering these three different categories would make deducing policy implications very complex and dependent on auxiliary assumptions.<sup>32</sup> It is beyond the scope of this paper to develop an optimal taxation model encompassing all of these categories. Instead, we will consider each of the findings in isolation and discuss their implications for optimal policies, other things equal.<sup>33</sup> Naturally, when determining actual governmental policies, many other things need to be taken into account, including collection and compliance costs, possibilities to allow for tax deductions, the convenience of introducing progression and special tax rates, and possibilities to impede evasion and avoidance. Hence, the following implications should be seen as an indication of the direction in which optimal policy may move.

### 6.1 Subjective Well-Being

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.<sup>34</sup> This finding thus suggests that, *ceteris paribus*, it would be better to levy taxes on the employer's side.

### 6.2 Labor Supply and Job Performance

Concerning labor supply, most economists would probably agree that taxes should be chosen that minimize the disruption of the price mechanism in the labor market. Be-

---

<sup>32</sup>For example, one needs to assume a specific social welfare function and decide what weight to attribute to preferences regarding public sector size and subjective well-being.

<sup>33</sup>The order in which we present the findings is from easiest to draw policy conclusions to least easy.

<sup>34</sup>In our experiment, we find large and highly significant differences in the case that tax proceeds are lost. We do not find any significant differences when taxes raised are used to produce a public good. As explained in the introduction, we view the real world as lying somewhere in between these two extremes. In this case, we expect subjective well-being to increase if more of the taxes are levied at the employers' side.

cause taxes are generally thought to reduce the efficiency of this market and to reduce labor supply, this implies that the tax should be chosen that maximizes labor supply (and output).<sup>35</sup> If people work more under one labor tax scheme than under the other, using this scheme brings society closer to first best and thus increases social welfare.

The experimental findings concerning labor supply are that people (especially men) tend to work more under an income tax than under an employer payroll tax.<sup>36</sup> Thus, the optimal policy for the government when considering only labor supply responses would be to rely only on income taxes.

### 6.3 Public Sector Size Preferences

It is less straightforward to draw policy conclusions from our results regarding preferences for the size of the public sector. To start, note that such preferences may be expressed in votes and therefore may affect the actual size of the public sector in the economy. If one assumes that there is something like an optimal size, it seems that the combination of labor taxes is optimal that induces individuals to prefer this size of the public sector. We cannot, however, draw conclusions from our finding that people prefer a larger public sector when the tax is levied on the employer's side as we cannot infer without additional assumptions what the 'objectively' optimal size of the public sector would be in the experiment.

There might thus be an optimal policy choice from a normative point of view. In a positive analysis, though, it is likely that a politician will try to use the distinct taxes to induce preferences in the electorate that are closer to her own view of optimal public sector size rather than to some 'objectively' optimal size.

---

<sup>35</sup>Assume that the government needs to raise a fixed amount of revenue to finance its expenditures and wants to do so via labor taxes. These taxes constitute a disincentive to work, such that labor supply is lower than under the first best solution, which consists of lump-sum taxes (assuming that labor supply elasticities are as usually reported (see Evers et al., 2008b), i.e. that the substitution effect is larger than the income effect).

<sup>36</sup>Similar evidence on labor supply is reported in Fochmann et al. (2010b) and Gamage et al. (2010).

## 7 Concluding Remarks

The question whether a labor tax levied at the employees' side of the labor market and one levied at the employers' side are equivalent is relevant for policy making, political economics and optimal taxation theory. In this paper we have investigated this LSE in a controlled laboratory experiment. Our results give support to 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 tends to be higher when the tax is levied on the employer's side, while labor supply and job performance tend to be lower. Furthermore, men and women respond differently to the framing of the tax.

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

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

## References

- Abeler, J., Falk, A., Götte, L., and Huffman, D. (2011). Reference points and effort provision. *The American Economic Review*, 101(2):470–492.
- Atkinson, A. and Stiglitz, J. (1980). *Lectures on public economics*. Economics handbook series. McGraw-Hill Book Co.

- Bassi, M. (2010). Mirrlees meets laibson: Optimal income taxation with bounded rationality. *CSEF Working Papers*.
- Becker, G. M., Degroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Systems Research and Behavioral Science*, 9:226–232.
- Blaufus, K., Bob, J., Hundsdorfer, J., Kiesewetter, D., and Weimann, J. (2010). It's all about tax rates: An empirical study of tax perception. *arqus Discussion Papers in Quantitative Tax Research*, (106).
- Blumkin, T., Ruffle, B. J., and Ganun, Y. (2012). Are income and consumption taxes ever really equivalent? evidence from a real-effort experiment with real goods. *European Economic Review*, 56(6):1200 – 1219.
- Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59.
- Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77.
- Conlisk, J. (1996). Why bounded rationality? *Journal of economic literature*, pages 669–700.
- Cox, J. C., Rider, M., and Sen, A. (2012). Tax incidence: Do institutions matter? an experimental study. Technical report, Experimental Economics Center, Andrew Young School of Policy Studies, Georgia State University.
- Crawford, V. P. and Meng, J. (2011). New york city cab drivers' labor supply revisited: Reference-dependent preferences with rationalexpectations targets for hours and income. *The American Economic Review*, 101(5):1912–1932.
- Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74.
- De Bartolome, C. A. M. (1995). Which tax rate do people use: Average or marginal? *Journal of Public Economics*, 56(1):79–96.
- Esarey, J., Salmon, T., and Barrilleaux, C. (2012a). Social insurance and income redistribution in a laboratory experiment. *Political Research Quarterly*, 65(3):685–698.

- Esarey, J., Salmon, T. C., and Barrileaux, C. (2012b). What motivates political preferences? self-interest, ideology, and fairness in a laboratory democracy. *Economic Inquiry*, 50(3):604–624.
- Evers, M., De Mooij, R., and Van Vuuren, D. (2008a). The wage elasticity of labour supply: a synthesis of empirical estimates. *De Economist*, 156(1):25–43.
- Evers, M., De Mooij, R., and Van Vuuren, D. (2008b). The wage elasticity of labour supply: A synthesis of empirical estimates. *De Economist*, 156:25–43.
- Fehr, E. and Goette, L. (2007). Do workers work more if wages are high? evidence from a randomized field experiment. *American Economic Review*, 97(1):298–317.
- Feldstein, M. (1999). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics*, 81(4):674–680.
- Fochmann, M., Kiesewetter, D., Blaufus, K., Hundsdorfer, J., and Weimann, J. (2010a). Tax Perception - An Empirical Survey. *Arqus Discussion Paper No. 99*.
- Fochmann, M., Kiesewetter, D., and Sadrieh, A. (2012). Investment behavior and the biased perception of limited loss deduction in income taxation. *Journal of Economic Behavior and Organization*, 81(1):230 – 242.
- Fochmann, M. and Weimann, J. (2011). The effects of tax salience and tax experience on individual work efforts in a framed field experiment.
- Fochmann, M., Weimann, J., Blaufus, K., Hundsdorfer, J., and Kiesewetter, D. (2010b). Grosswage illusion in a real effort experiment. *FEMM Working Papers*.
- Gamage, D. S., Hayashi, A., and Nakamura, B. K. (2010). Experimental evidence of tax framing effects on the work/leisure decision. *SSRN eLibrary*.
- Goette, L., Huffman, D., and Fehr, E. (2004). Loss aversion and labor supply. *Journal of the European Economic Association*, 2(2-3):216–228.
- Holm, P., Honkapohja, S., and Koskela, E. (1994). A monopoly-union model of wage determination with capital and taxes: An empirical application to the finnish manufacturing. *European Economic Review*, 38(2):285–303.
- Irtel, H. (2007). Pxlabs: The psychological experiments laboratory [online]. version 2.1.11. mannheim (germany): University of mannheim. [cited 25 april 2011]. available from <<http://www.pxlabs.de>>.

- Kahneman, D. and Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *The journal of economic perspectives*, 20(1):3–24.
- Kerschbamer, R. and Kirchsteiger, G. (2000). Theoretically robust but empirically invalid? an experimental investigation into tax equivalence. *Economic Theory*, 16:719–734.
- Koskela, E. and Schöb, R. (1999). Does the composition of wage and payroll taxes matter under nash bargaining? *Economics Letters*, 64(3):343–349.
- Lang, P. J. (1985). *The Cognitive Psychophysiology of Emotion: Anxiety and Anxiety Disorders*. Lawrence Erlbaum, Hillsdale, NJ.
- Lehmann, E., Marical, F. c., and Rioux, L. (2011). Labor earnings respond differently to income-tax and to payroll-tax reforms. IZA Discussion Papers 6108, Institute for the Study of Labor (IZA).
- McCaffery, E. and Slemrod, J. (2006). *Behavioral Public Finance*. G - Reference, Information and Interdisciplinary Subjects Series. Russell Sage.
- McCaffery, E. J. (1994). Cognitive theory and tax. *UCLA Law Review*, 41:1861–1947.
- McCaffery, E. J. and Baron, J. (2006). Thinking About Tax. *Psychology, Public Policy, and Law*, 12(1):106–135.
- Mullainathan, S., Schwartzstein, J., and Congdon, W. J. (2012). A reduced-form approach to behavioral public finance. *Annu. Rev. Econ.*, 4(1):511–540.
- Muysken, J., Van Veen, T., and ERIK, D. R. (1999). Does a shift in the tax burden create employment? *Applied Economics*, 31(10):1195–1205.
- O’Donoghue, T. and Rabin, M. (2003). Studying optimal paternalism, illustrated by a model of sin taxes. *American Economic Review*, 93(2):186–191.
- Oswald, A. J. (1985). The economic theory of trade unions: An introductory survey. *The Scandinavian Journal of Economics*, pages 160–193.
- Picard, P. and Toulemonde, E. (2001). On the equivalence of taxes paid by employers and employees. *Scottish Journal of Political Economy*, 48(4):461–470.
- Pissarides, C. A. (2000). *Equilibrium unemployment theory*. MIT press.
- R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

- Rasmussen, B. (1997). Non-equivalence of employment and payroll taxes in imperfectly competitive labor markets. *Univ. of Aarhus Dept. of Economics WP*, 22.
- Rasmussen, B. S. (1998). Long run effects of employment and payroll taxes in an efficiency wage model\*. *Economics Letters*, 58(2):245–253.
- Reimers, S. (2009). A paycheck half-empty or half-full? framing, fairness and progressive taxation. *Judgment and Decision Making*, 4(6):461–466.
- Riedl, A. and Tyran, J.-R. (2005). Tax liability side equivalence in gift-exchange labor markets. *Journal of Public Economics*, 89(11-12):2369–2382.
- Robins, P. K. (1985). A comparison of the labor supply findings from the four negative income tax experiments. *Journal of Human Resources*, pages 567–582.
- Sausgruber, R. and Tyran, J.-R. (2005). Testing the mill hypothesis of fiscal illusion. *Public Choice*, 122:39–68. 10.1007/s11127-005-3992-4.
- Sausgruber, R. and Tyran, J.-R. (2011). Are we taxing ourselves?: How deliberation and experience shape voting on taxes. *Journal of Public Economics*, 95(1-2):164 – 176.
- Shapiro, C. and Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *The American Economic Review*, pages 433–444.
- Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 69(1):99–118.
- Stiglitz, J. (2000). *Economics of the Public Sector*. Norton.
- Ullmann, R. and Watrin, C. (2008). Comparing direct and indirect taxation: The influence of framing on tax compliance. *European Journal of Comparative Economics*, 5(1):23–56.
- Weber, M. (2011). Income tax or employer’s contribution - an experiment on the perception of ‘economically equivalent’ duties. Mphil-thesis, Tinbergen Institute.



# A Appendix – Instructions, Test Questions and Questionnaire

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

## A.1 Instructions and Test Questions

FIRST INSTRUCTION SCREEN, ALL TREATMENTS

### **Welcome to this experiment!**

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

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

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

Please read all instructions carefully. You are not allowed to speak with other participants or to

communicate with them in any other way.

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

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

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

## SECOND INSTRUCTION SCREEN, ALL TREATMENTS

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

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

### **What the employees have to do:**

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

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

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

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

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

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

## FIRST PART OF THE THIRD INSTRUCTION SCREEN

### TREATMENT *EN*

#### **Payments:**

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

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

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

### TREATMENT *IN*

#### **Payments:**

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

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

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

### TREATMENT *EP*

**Payments:**

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

The wage an employee receives for a correct addition depends on how many problems this employee has already attempted to solve in this round. If the first attempt is correct, the employee's wage is 239.32 points. For each subsequent attempt the wage decreases by 19.94 points, but it is never smaller than zero. As employee you will always see on the screen the wage you will receive if the next addition is correct. For an incorrect addition neither the employer nor the employee will receive any points. The amount the employer receives for a correct addition of an employee is such that he/she will in the end always keep as a profit 49.8 points per correct addition.

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

**TREATMENT *IP*****Payments:**

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

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

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

## SECOND PART OF THE THIRD INSTRUCTION SCREEN, ALL TREATMENTS

### **Before the rounds start:**

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

If the random process we use determines that the extra round will in fact take place, the following occurs. The price of participating in this extra round is randomly determined and lies between 1056 and 2400 points. If this price is lower than the number you state, the price will be deducted from your earnings and you will work and get paid for an extra round. If the price is higher than the number you state, nothing will be deducted from your earnings and you will not work for an extra round. Note that we will not pay anyone until the whole experiment is finished. Note that you will have to state an amount of at least 1056 points (this is the amount of points you get for a round if you immediately switch to the fixed payment option).

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

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

## FIRST TEST QUESTION SCREEN, ALL TREATMENTS

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

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

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

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

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

How many minutes does each of the rounds last?

○

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

○

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

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

FIRST PART OF THE SECOND TEST QUESTION SCREEN

TREATMENT *EN*

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

- 33.3
- 66.7
- 50

TREATMENT *IN*

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

- 22
- 40
- 50

TREATMENT *EP*

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

- 33.3
- 66.7
- 50

#### TREATMENT *IP*

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

- 22
- 40
- 50

#### SECOND PART OF THE SECOND TEST QUESTION SCREEN, ALL TREATMENTS

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

- 

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

- 

#### THIRD TEST QUESTION SCREEN, ALL TREATMENTS

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

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

- 1300 points are deducted from your earnings, but you will not participate in the extra round.
- 1344 points are deducted from your earnings and you will participate in the extra round, if it takes place.
- 1300 points are deducted from your earnings and you will participate in the extra round, if it takes place.

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

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

- Yes.  
 No.

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

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

## SUMMARY OF THE INSTRUCTIONS FOR THE FIRST PART, ALL TREATMENTS

### **Summary of instructions**

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

You will be randomized into employees and employers.

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

If you are employer you hardly have to do anything.

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

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

## INSTRUCTIONS AFTER THE REGULAR WORK ROUNDS

### TREATMENTS *EN* AND *IN*

### **Welcome to the next part of the experiment**



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

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

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

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

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

TREATMENTS *EP* AND *IP*

### **Welcome to the next part of the experiment**

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

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

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

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

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

## INSTRUCTIONS BEFORE THE PREFERENCE FOR PUBLIC SECTOR SIZE ELICITATION

### TREATMENT *EN*

#### **Welcome to the next part of the experiment**

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

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

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

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

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

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

### TREATMENT *IN*

#### **Welcome to the next part of the experiment**

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

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

Some of the parameters will now be decided on by the employees. You will be asked to state which set of parameters you would prefer for this extra round of the experiment. Later, the

preferred set of parameters of one employee of your group will be chosen randomly and the additional extra round will take place with this set of parameters.

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

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

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

TREATMENT *EP*

### **Welcome to the next part of the experiment**

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

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

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

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

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

TREATMENT *IP*

## Welcome to the next part of the experiment

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

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

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

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

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

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

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

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

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

o

### 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 – Screenshots

Figure 5 shows a screenshot of an employee's computer during the work task (taken in treatment *IN*). Figure 6 shows the subjective well-being self-assessment using the SAM-V-9 self-assessment manikin. Figure 7 shows the slider that is used for the elicitation of preferences concerning the size of the public sector (taken in treatment *EN*). Figure 8 shows a screenshot of the loss aversion test.



Time: 06:53

Round 1 of 4

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

Figure 5: Screenshot during a work round

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

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 38 | 39 | 24 | 30 | 58 | 67 | 53 | 60 | 53 | 58 |
| 73 | 20 | 12 | 79 | 56 | 32 | 29 | 24 | 79 | 40 |
| 24 | 30 | 43 | 25 | 21 | 59 | 25 | 73 | 78 | 41 |
| 67 | 23 | 10 | 77 | 35 | 30 | 15 | 25 | 40 | 76 |
| 48 | 58 | 20 | 36 | 28 | 35 | 15 | 29 | 49 | 57 |
| 33 | 71 | 30 | 53 | 44 | 72 | 65 | 55 | 56 | 49 |
| 73 | 62 | 39 | 50 | 78 | 35 | 72 | 22 | 69 | 44 |
| 54 | 14 | 71 | 50 | 63 | 42 | 71 | 52 | 70 | 17 |
| 25 | 12 | 13 | 23 | 62 | 68 | 71 | 41 | 65 | 43 |
| 42 | 16 | 24 | 74 | 38 | 68 | 32 | 56 | 65 | 74 |

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 83 | 57 | 25 | 13 | 51 | 34 | 66 | 42 | 74 | 63 |
| 12 | 69 | 29 | 28 | 19 | 37 | 53 | 77 | 57 | 74 |
| 22 | 22 | 58 | 46 | 75 | 82 | 66 | 82 | 13 | 70 |
| 14 | 50 | 47 | 58 | 21 | 76 | 56 | 81 | 19 | 30 |
| 27 | 62 | 52 | 38 | 59 | 36 | 54 | 42 | 68 | 74 |
| 67 | 21 | 78 | 64 | 25 | 67 | 78 | 77 | 46 | 83 |
| 20 | 73 | 63 | 13 | 40 | 51 | 45 | 57 | 15 | 45 |
| 74 | 83 | 81 | 52 | 46 | 71 | 43 | 34 | 42 | 37 |
| 37 | 62 | 31 | 34 | 13 | 23 | 20 | 78 | 36 | 62 |
| 54 | 75 | 55 | 67 | 65 | 40 | 59 | 34 | 23 | 28 |

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

Go to fixed payment option

Figure 6: Screenshot during the self-assessment with the SAM-V-9

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

The image shows a self-assessment scale for mood. It consists of nine faces, numbered 1 to 9, arranged in a horizontal row. Each face has a different expression, ranging from a neutral, slightly smiling face (1) to a very sad, crying face (9). Below the faces is a legend with nine radio buttons, each corresponding to a number from 1 to 9. To the right of the legend is an 'OK' button.

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

OK

Figure 7: Screenshot of the slider; public sector size preference

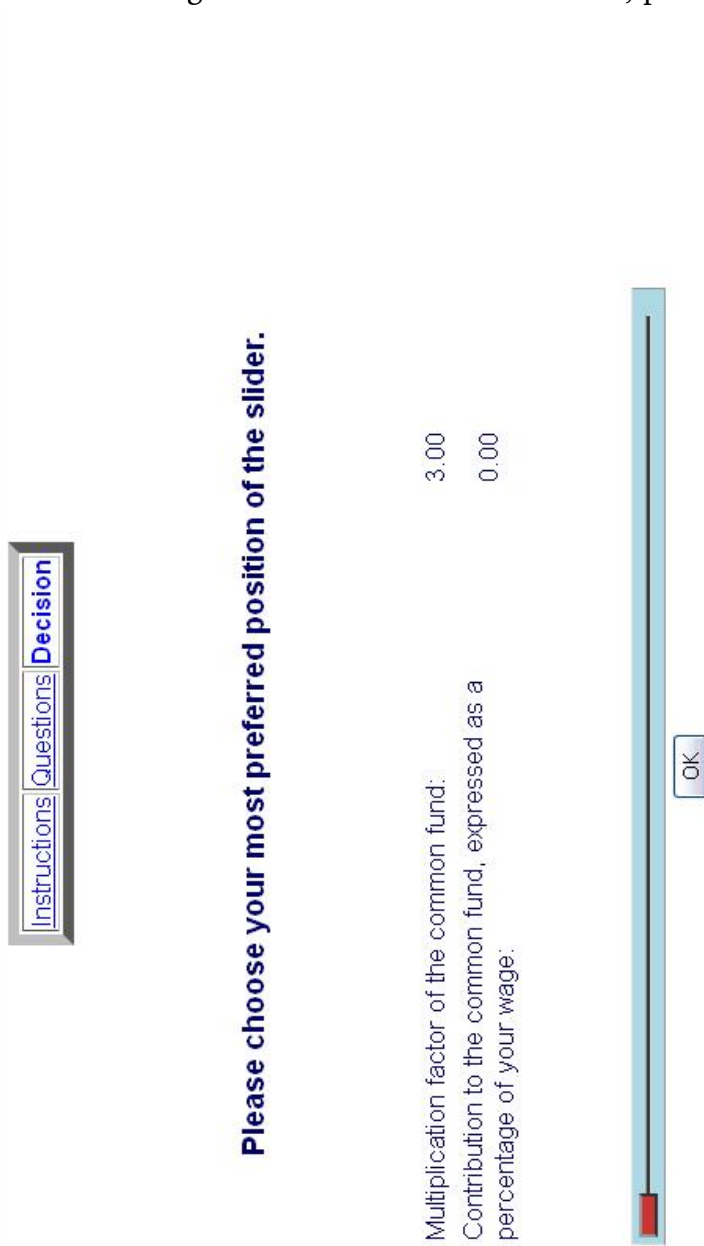


Figure 8: Screenshot while subjects choose which lotteries they want to participate in

