Essays in Behavioral Economics

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von
Johannes Abeler
aus Münster/Westf.

Bonn 2008

Dekan:	Prof. Dr. Erik Theissen
Erstreferent:	Prof. Dr. Armin Falk
Zweitreferent:	Prof. Paul Heidhues, Ph.D.

Tag der mündlichen Prüfung: 03.07.2008

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn (http://hss.ulb.uni-bonn.de/diss_online) elektronisch publiziert.

Acknowledgments

First and foremost, I am grateful to Armin Falk for his advice during many discussions and for his unconditional support of my research projects. His way of doing research, defining relevant research questions, and tackling them in an innovative way, has been exemplary for me.

Paul Heidhues provided very helpful comments and support to this dissertation in all stages. Jörg Oechssler was always ready to listen to my ideas and to give valuable advice.

I want to thank my co-authors Sebastian Kube, Lorenz Götte, and David Huffman, and especially Steffen Altmann, Felix Marklein, and Matthias Wibral. It was fun working, discussing, rewriting the paper, and rewriting the paper with them.

During most of my dissertation, I enjoyed the hospitality of the Institute for the Future of Labor (IZA). I learned a lot from resident and visiting researchers. I thank all of them and additionally Iwona Werner for doing a superb organizing job.

Financial support from the DFG (through GRK 629 and SFB TR/15) and from the Bonn Graduate School of Economics is gratefully acknowledged. I especially thank Urs Schweizer and Georg Nöldeke for their continuing efforts in organizing the BGSE and facilitating research as much as possible.

My fellow graduate students, especially Philipp Kircher, Sebastian Kranz, Susanne Ohlendorf, and David Schröder, made Bonn an exciting and genial place to do economic research.

Finally, I want to thank my parents who supported me in every aspect of what I did and in particular Lucie for her encouragement, care, and endurance. Her non-economist views brought my research closer to the *real world*.

Contents

List of Figures

List of Tables

Introduction

The roots of behavioral economics date back to Adam Smith, who described human decision making in his book "The Theory of Moral Sentiments" as driven by passions, myopia, and moral concerns (Ashraf et al. 2005). In his second book "The Wealth of Nations", man had transformed into a rational being, selfishly maximizing his own monetary gain. Economics followed the second path and developed a powerful toolbox to predict human behavior resting on the assumptions of full rationality and self-interest. This neo-classical model of utility maximization is tremendously successful in describing and predicting economic behavior. However, many relevant economic phenomena are still hard to capture in this framework. Therefore, beginning with Simon's (1955) pioneering analysis of bounded rationality, behavioral aspects came again into the focus of economic research. The break-through for behavioral economics was Kahneman & Tversky's (1979) seminal study on non-standard preferences. Central building blocks of their prospect theory are reference dependence and loss aversion, i.e., that people compare their outcome with a reference point and that, in this comparison, losses loom larger than equal-sized gains. Kahneman and Tversky's study has initiated a tremendous amount of research on how much human decision making is in line with the predictions of the neo-classical model.

Behavioral economics has investigated questions related to many field in economics (for an overview, see Camerer 2002). In labor and personnel economics, for example, it has been found that social preferences, such as reciprocity (Rabin 1993, Falk & Fischbacher 2006), inequity aversion (Fehr & Schmidt 1999, Bolton & Ockenfels 2000) and altruism (Andreoni & Miller 2002) are important driving forces for decisions and that people do not only care for their *own* economic well-being.

In the area of saving, investment, and consumption, models of temporal decision-making and limited self-control (e.g., Laibson 1997, Fudenberg & Levine 2006) have substantially influenced economic research, demonstrating that many people often act in a time-inconsistent way, overweighing the present and neglecting the distant future. One phenomenon that is implicitly assumed in many behavioral models and is also of direct importance is "narrow bracketing," i.e., that people do not decide globally but rather evaluate parts of a decision separately (Thaler 1985, Read et al. 1999*a*). Many decisions people face in life, especially concerning investment and consumption, are very complex. All aspects of the decisions have to be taken into account at the same time to find an optimal solution. Bracketing narrowly thus facilitates decision-making but prevents people from finding the global optimum and makes them vulnerable to framing effects (Kahneman & Lovallo 1993).

This dissertation aims at continuing and expanding three lines of research prominent in behavioral economics today. Chapter 1 deals with the role of reciprocity in labor contracts and explores which factors determine whether a payment scheme is perceived as fair. Chapter 2 studies whether people integrate money from various income sources. This question is related to the notion of narrow bracketing. Chapter 3 tests whether people evaluate their outcome in comparison to a reference point. More specifically, it tests whether an individual's rational expectations serve as a reference point.

Even though the three chapters cover different areas of behavioral economics, they share the same underlying questions: To which degree are fundamental assumptions of economics in line with real world behavior and where do they need to be refined? And how can we use this new knowledge to design economic institutions serving people that are neither fully rational nor completely selfish? All chapters present empirical analyses of laboratory or field experiments. Experiments provide a maximum of control and allow for a causal interpretation as treatments are assigned exogenously. The results presented in this dissertation are potentially important for many fundamental economic questions, especially in the areas of labor and public economics, but also for the understanding of investment and consumption decisions.

Chapter 1 focuses on how payment schemes should be designed if workers are reciprocal. In recent years, a growing literature has shown that reciprocity, i.e., the willingness to incur personal costs in order to reward kind actions or punish unkind ones, can mitigate the enforcement problems of incomplete contracts (e.g., Fehr & Gächter 1998, Fehr & Gächter 2000, Fehr & Schmidt 2003). The potential of reciprocity as a contract enforcement device, however, is likely to depend on the institutions that shape the employment relation, above all the mode of payment. Yet little is known about the interaction of reciprocity with different payment modes. Exploring this interaction is crucial in order to understand under which pay scheme the efficiency-enhancing effects of reciprocity develop their full power.

In this chapter, we study the most common wage institution, namely horizontal wage equality. In particular we address the following questions: Is wage equality an efficient payment scheme when contract enforcement relies on reciprocity? Do principals choose to pay equal wages if they are not constrained to? How does the mode of payment affect work morale? Are equal wages perceived as fairer compared to individual wages? Paying equal wages for equal performance seems uncontroversial in this respect. But in real-life work relations this case is likely to be the exception rather than the rule. When workers differ in their performance, reciprocity probably implies that the agent who works more should also get a higher payoff compared to his co-worker. In other words, it might well be that the often-heard slogan "equal pay for equal work" also calls for "unequal pay for unequal work".

We study these questions in a simple and parsimonious laboratory experiment. In the experiment, one principal is matched with two agents. In a first stage the agents exert costly effort. After observing their efforts, the principal pays them a wage. In the main treatment he can choose the level of the wage but he is obliged to pay the same wage to both agents. In the control treatment, the principal can wage discriminate between the two agents. In both treatments, neither efforts nor wages are contractible.

The main findings of the experiment are as follows. First, performance differs substantially between the two treatments: agents who are paid equal wages exert significantly lower efforts than agents who are paid individually. Effort levels are nearly twice as high under individual wages. In addition, efforts decline over time when equal wages are paid. Second, this strong treatment effect cannot be explained by differences in monetary incentives. The actual wage choices of principals imply that providing high effort levels is profitable for agents in both treatments. From a purely monetary viewpoint agents' behavior in both treatments should thus be similar. Third, we show that the frequent violation of the norm of reciprocity in the equal wage treatment can explain the effort differences between the treatments. In both treatments, agents who exert a higher effort and earn a lower payoff than their co-worker strongly decrease their effort in the next period. This pattern is very similar in both treatments. However, the norm of reciprocity is violated much more frequently under equal wages.

Regarding the design of institutions in firms, our findings highlight the importance of taking the concerns for co-workers' wages into account. However, doing so by paying equal wages to a group of agents may actually do more harm than good. As soon as agents differ in their performance, equal wages might be considered very unfair.

Chapter 2 combines both fundamental questions addressed in this dissertation: it tests a central economic assumption, namely whether people integrate different income sources; and it develops how the design of institutions influences people who do or do not integrate their income sources. Usually, economic theory assumes that people don't care where their money came from, or in more technical terms, assumes that money is fungible. This is a fundamental assumption in economics, but there is only very little controlled evidence for or against it. We pursue a dual research strategy by combining a field experiment and a laboratory experiment.

Both experiments have the same setup: subjects receive a subsidy in addition to their normal budget. In the main treatment, the subsidy is paid in kind, i.e., it has to be spent on one specific good. Subjects in a control treatment receive the subsidy in cash. The crucial detail of the experimental setup is that the amount of the in-kind subsidy is smaller than the amount optimally spent on the subsidized good. Giving the subsidy in kind or in cash should thus not distort consumption if subjects integrate their income sources: by shifting the remaining budget, subjects can reach the same first-best consumption level in both treatments. In contrast, subject who do not treat different income sources as fungible will spend the in-kind subsidy fully on the subsidized good and thus consume too much of this good.

The field experiment was conducted in a wine restaurant. We chose this setting because here almost all guests consume two goods: they eat and they drink. In one week, guests got a "beverage voucher" that had to be spent on beverages. The value of the voucher was lower than the amount that guests spend anyway on beverages. In the next week, guests got a voucher for the total bill. Guests did not know that they participated in an experiment, and thus acted in a completely natural environment without feeling observed. We find that giving an 8-euro beverage voucher increases beverage consumption by 3.90 euro compared to the control treatment. This additional spending does not lead to higher beverage consumption in terms of volume; instead guests buy more expensive bottles of wine. We have also data on guests who visited the restaurant before and after the experiment. Compared to these guests, guests who got either voucher spend 3.93 euro more on total.

Our second experiment was conducted in the laboratory. This setup allows us to control the environment even more than in the field experiment. We induce the payoff function and we endow every participant with the same budget. We can thus guarantee that the subsidy is inframarginal for every subject, something which we cannot do in the field experiment. In the lab, we can also elicit additional information about subjects. The laboratory experiment confirms the main result of the field experiment: subjects spend significantly more on the subsidized good when it is given in kind. Subjects spend 57 percent of the subsidy on the subsidized good when they receive the in-kind subsidy, while for subjects in the control treatment this figure is only 28 percent. We can thus double the marginal propensity to consume the subsidized good just by attaching a label to the subsidy. Last but not least, we find that subjects with lower mathematical ability are more likely to be influenced by our treatment variation.

The specific design of our experiment points to a new application in public economics, as it gives valuable insights into the effect that in-kind benefit programs have on market outcomes. Recently, the economic literature has documented a strong impact of housing benefits on market rents for several countries (Susin 2002, Gibbons & Manning 2006, Fack 2006). For the U.S., where not all eligible households do receive the benefits, Susin (2002) finds that housing benefits even caused a *net loss* for low-income households. According to our results, tenants who receive housing benefits will have a higher willingness to pay for a given apartment than tenants who get the same amount as a cash grant. If landlords realize this behavior they could seize an increase in housing benefits to increase the rent. Taking this effect into account when designing in-kind benefit programs would at the same time help low-income households and save tax payers' money. This problem could be mitigated, for example, by linking housing benefits less saliently to rent payments. The periodicity of the benefit payments could be chosen such that it differs from the periodicity of the rent payments. Moreover, one could design the benefit system such that the exact amount of the subsidy depends on variables which the landlord cannot observe.

In Chapter 3 we test whether the rational expectations of a person serve as a reference point. Theories of reference-dependent preferences predict that individuals value changes in income relative to a reference point, rather than simply the level of income. A key open question, however, is what people take as the reference point for their decisions. The original formulation of reference dependent preferences by Kahneman & Tversky (1979) is remarkably silent on what constitutes a reference point. Most studies so far have taken the status quo or lagged status quo to be the reference point (e.g., Odean 1998, Genesove & Mayer 2001). A more recent class of models assumes that reference points are expectations rather than the status quo (Bell 1985, Loomes & Sugden 1986, Gul 1991, Kőszegi & Rabin 2006, Kőszegi & Rabin 2007). This distinction is relevant in the many situations in life where expectations are different from the status quo. For example, a salary increase of 20,000 would be a gain relative to the status quo, but could be a loss if the expected increase was 50,000, with strongly different implications for behavior depending on what the reference point is. Other examples where expectations intuitively serve as a reference point include deciding whether to buy a good with an unexpectedly high price (Heidhues & Kőszegi forthcoming), not getting an item on eBay after having been the highest bidder, or being rejected after a "revise and resubmit" compared to being rejected directly.

This chapter reports evidence from an experiment in which expectations are distinct from the status quo, and where we exogenously vary expectations to test whether this affects behavior in the form of real effort choices (labor supply). A key advantage of the experiment is that the rational expectations of participants are known to the researcher, something which is difficult to achieve in the field. In the experiment, subjects work on a tedious, repetitive task and can decide after each repetition whether to continue or to stop working. They get a piece rate, but receive their accumulated piece rate earnings only with 50 percent probability. With 50 percent probability they receive a fixed payment instead. Which payment subjects receive is determined only after they have made their choice about when to stop working. The only treatment manipulation involves varying the size of the fixed payment.

Standard theory and status-quo reference dependence both predict no treatment difference in the experiment. In the standard model with separable utility over money and effort costs, the optimal effort level is determined by setting marginal cost equal to the marginal benefit defined by the piece rate; the fixed payment is irrelevant for either marginal cost or benefits of effort. This is true independent of the shape of utility over money, and the shape of the cost function. Models where the status quo is the reference point also predict no treatment difference, because the status quo when entering the experiment is the same for both treatments. A model with expectations-based reference points such as Kőszegi & Rabin (2007) predicts higher effort in the treatment with the higher fixed payment. As long as the accumulated piece rate is below the fixed payment, the individual risks getting the accumulated piece rate instead of the (higher) fixed payment, which would feel like a loss. This expectation induces the worker to exert more effort. Once accumulated earnings are higher than the fixed payment, however, getting the fixed payment feels like a loss, and working harder only magnifies the size of the loss. This expectation induces lower effort. The model thus predicts that individuals will tend to choose accumulated earnings that exactly equal the size of the fixed payment, and that increasing the size of the fixed payment will tend to increase overall effort.

We find a strong treatment effect, such that individuals in the high fixed payment treatment work significantly more. We also observe strong spikes in the distribution of effort choices, exactly at the respective fixed payment amount in each treatment. We also find support for a more subtle prediction, which is that the treatment difference should be driven by individuals who stop in the interval between the low and high fixed payment amounts.

Besides a contribution to the fundamental question what constitutes a reference point, our results are also relevant for the lively debate regarding labor supply and transitory wage changes. It is debated whether the response of effort to changes in incentives is consistent with the standard intertemporal substitution of labor and leisure or rather with loss aversion around a daily reference income (e.g., Camerer et al. 1997, Farber 2005, Fehr & Götte 2007). A key lacuna in this literature is that the reference point is unobserved; our experiment makes the reference point known to the researcher. As noted by Kőszegi & Rabin (2007), if reference points are based on expectations, the labor supply reaction to incentives depends on whether the change in incentives is anticipated or not. For example, if an individual expects the hourly wage to be low on a given day, earning a small amount does not feel like a loss. But if the hourly wage is *unexpectedly* low, this does feel like a loss relative to expectation, and can induce workers to work even harder to try to reach their expectation, contrary to the standard prediction on intertemporal substitution. This distinction potentially helps reconcile some of the apparently conflicting findings in the field evidence, as in Camerer et al. (1997) and Farber (2005).[1]

[1]Chapter 1 was developed in collaboration with Steffen Altmann, Sebastian Kube, and Matthias Wibral and was circulated under the title "Reciprocity and Payment Schemes: When Equality Is Unfair". The laboratory experiment in Chapter 2 is joint work with Felix Marklein and the chapter was circulated under the title "Fungibility, Labels, and Consumption". Chapter 3 owes to the collaboration with Armin Falk, Lorenz Götte, and David Huffman.

Chapter 1

When Equality is Unfair

"To treat people fairly you have to treat people differently."
Roy Roberts, at that time VP of General Motors[1]

1.1 Introduction

In recent years, a vast body of literature has stressed the influence of social norms on individual decision making processes. Especially reciprocity, i.e., the willingness to reward kind actions and punish unkind actions even at a cost, has proven to be a highly relevant norm (e.g., Fehr & Gächter 1998, Fehr & Gächter 2000, Fehr & Schmidt 2003). In employment relations, reciprocity can mitigate the enforcement problems of incomplete contracts: many agents repay a gift in the form of higher wages by providing higher efforts even in one-shot situations where no future gains can be expected (e.g., Fehr et al. 1997, Hannan et al. 2002, Fehr & Falk 2002, Maximiano et al. forthcoming)). This "gift exchange" (Akerlof 1982) constitutes at least a partial solution to moral-hazard problems that are widespread in labor relations.

The potential of reciprocity as a contract enforcement device, however, is likely to depend on the institutions that shape the employment relation, above all the mode of payment. Yet little is known about the interaction of reciprocity with different

[1] Quoted in (Baker et al. 1988).

payment modes. Exploring this interaction is crucial in order to understand under which pay scheme the efficiency-enhancing effects of reciprocity develop their full power. This is what this chapter does. The specific wage institution we study is wage equality. In particular we address the following questions: Is wage equality an efficient payment scheme when contract enforcement relies on reciprocity? Do principals choose to pay equal wages if they are not constrained to? How does the mode of payment affect work morale? Are equal wages perceived as fairer compared to individual wages?

Paying equal wages to workers on the same level of a hierarchy is common practice in many firms (e.g., Medoff & Abraham 1980, Baker et al. 1988). Several reasons for equal wages have been brought forward, amongst them increased peer monitoring (Knez & Simester 2001) and lower transaction costs since contracts do not have to be negotiated with every worker individually (see also Prendergast 1999). In addition, a concern for fairness has been a main argument invoked to justify equal wages. It has been argued that differential pay of co-workers is considered unfair by workers, causes resentment and envy within the workforce, and ultimately lower performance (Akerlof & Yellen 1990, Bewley 1999, Milgrom & Roberts 1992). Equality is also often referred to in employer-union bargaining as being a cornerstone of a fair wage scheme.

We contribute to the discussion about equal wages by examining their impact on the effectiveness of reciprocity in enforcing incomplete contracts. Paying equal wages for equal performance seems uncontroversial in this respect. But in real-life work relations this case is likely to be the exception rather than the rule. When workers differ in their performance the following questions need to be answered. Does reciprocity imply that the agent who works more should also get a higher payoff compared to his co-worker? In other words, does the often-heard slogan "equal pay for equal work" also call for "unequal pay for unequal work"? If this is the case, does a high-performing agent become frustrated and decrease work effort under equal wages?[2] How do low-performing agents react? What are the

[2]Lazear (1989) raises similar doubts about pay equality (p. 561): "It is common for both management and worker groups such as labor unions to express a desire for homogeneous wage

consequences for efficiency?

Ideally, these questions would be examined in two work environments that differ only with respect to the payment mode. To come close to this ideal world, we introduce a simple and parsimonious laboratory experiment that allows us to analyze the interaction between the institution of wage equality and reciprocity. In the experiment, one principal is matched with two agents. In a first stage the agents exert costly effort. After observing their efforts, the principal pays them a wage. In the main treatment he can choose the level of the wage but he is obliged to pay the same wage to both agents (*equal wage treatment* or EWT). In the control treatment, the principal can wage discriminate between the two agents (*individual wage treatment* or IWT). In both treatments, neither efforts nor wages are contractible. Note that principals in the individual wage treatment are free to pay the same wage to both agents, i.e., the EWT is a special case of the IWT.

The main findings of the experiment are as follows. First, performance differs substantially between the EWT and the IWT: agents who are paid equal wages exert significantly lower efforts than agents who are paid individually. Effort levels are nearly twice as high under individual wages. In addition, efforts decline over time when equal wages are paid. Second, this strong treatment effect cannot be explained by differences in monetary incentives. The actual wage choices of principals imply that providing high effort levels is profitable for agents in both treatments. From a purely monetary viewpoint agents' behavior in both treatments should thus be similar. Third, we show that the frequent violation of the norm of reciprocity in the equal wage treatment can explain the effort differences between the treatments. In both treatments, agents who exert a higher effort and earn a lower payoff than their co-worker strongly decrease their effort in the next period. This pattern is very similar in both treatments. However, the norm of reciprocity is violated much more frequently under equal wages. Principals in the IWT understand the mechanisms of reciprocity quite well. When efforts differ they do pay different wages, rewarding

treatment. The desire for similar treatment is frequently articulated as an attempt to preserve worker unity, to maintain good morale, and to create a cooperative work environment. But it is far from obvious that pay equality has these effects."

the harder-working agent with a higher payoff in most cases.

Agents' reactions cause completely different dynamics in the two treatments. Under equal wages, initially hard-working agents get discouraged and reduce their effort to the level of their low-performing co-workers. By contrast, in the individual wage treatment the high performers keep exerting high efforts while the low performers change their behavior and strongly increase their effort levels.

Our results suggest a psychological rationale for using individual wages. Subjects perceive equal wages for unequal performance as unfair and reduce their effort subsequently. The traditional literature on incentive provision in groups comes to a similar conclusion though for a different reason. It is usually argued that the inefficiency of equal wages stems from the fact that marginal products and wages are not aligned. This can lead to free-riding among selfish agents (e.g., Holmstrom 1982, Erev et al. 1993). We enlarge the scope of this critical view on wage equality: interestingly, in our setup it is precisely the presence of reciprocal agents and not their absence that calls for the use of individual rewards.

Since agents in our experiment compare their payoff with the payoff of their co-worker, our results also inform the literature analyzing the influence of relative income on satisfaction and performance. It has been shown that relative income affects people's well-being (e.g., Clark & Oswald 1996, Easterlin 2001). However, it is less clear how this influences performance, i.e., whether low relative income leads to frustration and reduced performance (e.g., Clark et al. 2006, Torgler et al. 2006) or to an increase in performance due to a "positional arms race" (Neumark & Postlewaite 1998, Layard 2005, Bowles & Park 2005). The controlled laboratory environment of our experiment allows us to reconcile these differing views. Our results indicate that the comparison process goes beyond a one-dimensional comparison of income and also includes a comparison of effort. In particular, they suggest that receiving a lower income while exerting a *higher* effort leads to reduced performance which is likely to be driven by feelings of exploitation. By contrast, a lower income that is generated by a *lower* effort leads to a (small) increase in performance.

There are only a few experimental studies that analyze the interaction of payment modes and social preferences when contracts are incomplete.[3] Fehr et al. (forthcoming) let principals choose between contracts relying on explicit incentives and "bonus contracts" relying on trust and reciprocity. In a bilateral setup, they find that a bonus contract oftentimes yields a higher efficiency than the incentive contract. Most closely related to this chapter is the work of Charness & Kuhn (2005). Here, one principal is matched with two agents differing in productivity; like in our study, wages and efforts are not contractible. In contrast to our results, they find that co-workers' wages do *not* matter much for agents' decisions. However, their design differs from ours in several important points. While Charness & Kuhn (2005) focus on heterogeneity in productivity, we look at the effect of actual output differences between agents. Furthermore, we allow for richer comparisons between the agents, as in their design agents are not aware of the magnitude and direction of the productivity differences. The different results underline the importance of information for determining the reference group: Charness & Kuhn (2005)'s results rather apply to groups of workers that are loosely related and know little about each other, while our focus is on close co-workers who have a good understanding about their peers' abilities and efforts.

Regarding compensation practice in firms, our findings highlight the importance of taking the concerns for co-workers' wages into account. However, doing so by paying equal wages to a group of agents may actually do more harm than good. As soon as agents differ in their performance, equal wages which seem to be a fair institution at first sight might be considered very unfair. While the discouraging effect of equal wages on hard-working agents has long been informally discussed (e.g., Milgrom & Roberts 1992) this chapter provides controlled evidence in favor of this intuition. Moreover, it suggests that it is the violation of reciprocity that causes the discouragement and low performance. Our results should not be interpreted as arguments against wage equality in general but they rather point to limits of equal wages. Wage equality is potentially a good choice in occupations where, e.g.,

[3] Theoretical analyses of this interaction are provided by, for example, Demougin & Fluet (2003), Bartling & von Siemens (2004), and Itoh (2004). All these models rely on purely distributive (outcome-based) preferences like envy or inequality aversion. For most cases they do not predict that equal wages do lead to lower effort exertion as it is the case in our experiment.

due to technological reasons, workers' performance differs only slightly or where performance differences are due to random influences. In addition, the transparency of co-workers' work efforts and wages might have an influence on the optimal choice of the pay scheme.

The remainder of this chapter is structured as follows. In the next section we describe the experimental design and discuss theoretical predictions. In section 1.3 we present and discuss results and section 1.4 concludes.

1.2 Experimental Setup

1.2.1 Experimental Design

In the experiment, one principal is matched with two agents. The subjects play a two-stage game. In the first stage, the agents decide simultaneously and independently how much effort they want to exert. Exerting effort is costly for the agents. Effort choices range from 1 to 10 and are associated with a convex cost function displayed in Table 1.1. The principal reaps the benefits of production: every unit of effort increases his payoff by 10.

Effort level e_i	1	2	3	4	5	6	7	8	9	10
Cost of effort $c(e_i)$	0	1	2	4	6	8	10	13	16	20

Table 1.1: *Cost of effort.*

In the second stage, after observing the effort decisions of the agents in his group, the principal decides on wages for the two agents. The wages have to be between 0 and 100. Neither efforts nor wages are contractible. The only difference between the treatments is the mode of payment. In our main treatment the principal can only choose one wage w that is paid to each of the agents (*equal wage treatment* or EWT). In the control treatment he can discriminate between the two agents by choosing wages w_1 and w_2 for agent 1 and 2, respectively (*individual wage treatment* or IWT). The EWT is thus a special case of the IWT. At the end of each period, the two agents and the principal are informed about efforts, wage(s), and the resulting payoffs for all three players. The payoff functions for the players are summarized in Table 1.2.

Treatment	EWT	IWT
Payoff Principal	$\pi_P = 10(e_1 + e_2) - 2w$	$\pi_P = 10(e_1 + e_2) - (w_1 + w_2)$
Payoff Agent i	$\pi_{A_i} = w - c(e_i)$	$\pi_{A_i} = w_i - c(e_i)$

Table 1.2: *Payoffs of players.*

This game is played for twelve periods. We implemented a stranger design to abstract from confounding reputation effects, i.e., at the beginning of each period principals and agents were rematched anonymously and randomly within a matching group. A matching group consisted of three principals and six agents. The subjects kept their roles throughout the entire experiment. After the last period, subjects answered a questionnaire. The experiment was conducted in a labor market framing, i.e., principals were called "employers" and agents were called "employees".[4]

Our setup is related to the gift-exchange game (Fehr et al. 1993) but differs in two important ways. First, a principal is matched with two agents instead of one. This is an essential prerequisite to analyze the interaction between the institution of wage equality and reciprocity. Additionally, the agents move first while in most experiments the principal moves first. Our move order allows the principal to base his wage decision on the actually exerted effort.[5]

All participants started the experiment with an initial endowment of 400 points that also served as their show-up fee. Points earned were converted at an exchange rate of 0.01 Euro/point. The experiment was conducted at the BonnEconLab at the University of Bonn in April 2005 using z-Tree (Fischbacher 2007). For each treatment, we ran four sessions with a total of 8 matching groups (144 participants). The experiment lasted approximately 70 minutes. On average subjects earned 8.30

[4]An English translation of the instructions is available from the authors upon request.

[5]For an experiment in which principals move first and decide according to productivity differences, see Charness & Kuhn (2005). See Gneezy (2006) for a direct comparison of move orders.

Euro including the show-up fee of 4 Euro.

1.2.2 Behavioral Predictions

Efficiency is determined by agents' effort choices. It is maximized if both agents exert the highest possible effort of 10. However, if all players are rational and selfish the principal will not pay anything to the agents since wage payments only reduce his monetary payoff. Anticipating this, both agents will provide the minimal effort of one in the first stage. The finite repetition of the game in randomly rematched groups does not change this prediction. This subgame perfect equilibrium is the same for both payment modes. If all players were selfish we should therefore expect no difference between treatments.

By contrast, in laboratory experiments studying labor relations with incomplete contracts, one typically observes that efforts and wages exceed the smallest possible value. Moreover, wages and efforts are positively correlated (e.g., Fehr & Gächter 2000). These stylized facts can be explained with a preference for reciprocity, i.e., players reward kind actions of other players and punish unkind actions, even if they have to incur a cost for the reward or punishment. This implies that a higher effort will be rewarded with a higher wage. More importantly, the two treatments of our experiment can yield different outcomes if players are reciprocal. In our three-player setup, reciprocity might imply that the agent who works more also should get a higher payoff compared to his co-worker.[6] This is not possible under wage equality when agents differ in their performance. Since agents are paid the same wage and have to bear the cost of effort exertion, the agent who exerted a higher effort receives a lower net payoff and feels exploited. Providing an additional unit of effort increases the risk of being exploited and therefore constitutes a disincentive to provide effort for reciprocal agents under equal wages.[7]

[6]Note that this concept of reciprocity does not solely rely on intentions of the principal but rather captures procedural fairness more generally (cf. Dufwenberg & Kirchsteiger (2004) or Falk & Fischbacher (2006) for formal models of reciprocity).

[7]The co-worker of an exploited agent may also experience some disutility. He works less but earns more than his colleague, so he could feel guilty about this undeserved profit. However, it is reasonable to assume that a disadvantageous norm violation (exploitation) is experienced more

On the contrary, in the individual wage treatment it is always possible for a principal to fulfill the norm of reciprocity. He can set the wages such that the hard-working agent gets a higher payoff. If principals mostly do so, one will observe fewer norm violations in the IWT than in the EWT. If agents react similarly to norm violations in both treatments, efforts in the EWT should be lower than under individual wages.[8]

1.3 Results

In this section we present the results of the experiment and discuss possible explanations for the observed behavior. We first analyze the efficiency implications of the two payment schemes by comparing the effort choices of agents. We then explore possible reasons for the effort choices by analyzing how the mode of pay affects monetary incentives and interacts with the social preferences of the agents.

1.3.1 Effort Choices and Efficiency

Figure 1.1 shows the development of average efforts over time. Two things are striking about the graph. First, efforts are considerably lower in the equal wage treatment. While agents in the IWT on average exert an effort of 8.21, agents in the EWT only provide an effort of 4.40 (Mann-Whitney test: $p < 0.01$).[9] Second, efforts decrease over time under equal wages which is not the case when individual wages are paid (Wilcoxon test for periods 1–6 against 7–12: IWT, $p = 0.56$; EWT,

strongly than an advantageous norm violation (guilt). Support for this assumption can be found in, e.g., Loewenstein et al. (1989) and Babcock et al. (1996).

[8]Other models of social preferences assume that players do not care about reciprocity per se but dislike unequal payoffs, e.g., the model by Fehr & Schmidt (1999). Using their preferred parameters, their model predicts the same outcome in our game as the subgame perfect Nash equilibrium: minimal wages ($w = 0$) and minimal efforts ($e = 1$). Assuming extreme values for the guilt parameter ($\beta > 2/3$ for 40% of subjects) does not change this result by much. In this case, the model predicts an average effort of 1.6. This prediction is the same for both treatments. Calculations are available upon request.

[9]Unless otherwise noted, all tests use matching group averages as independent observations.

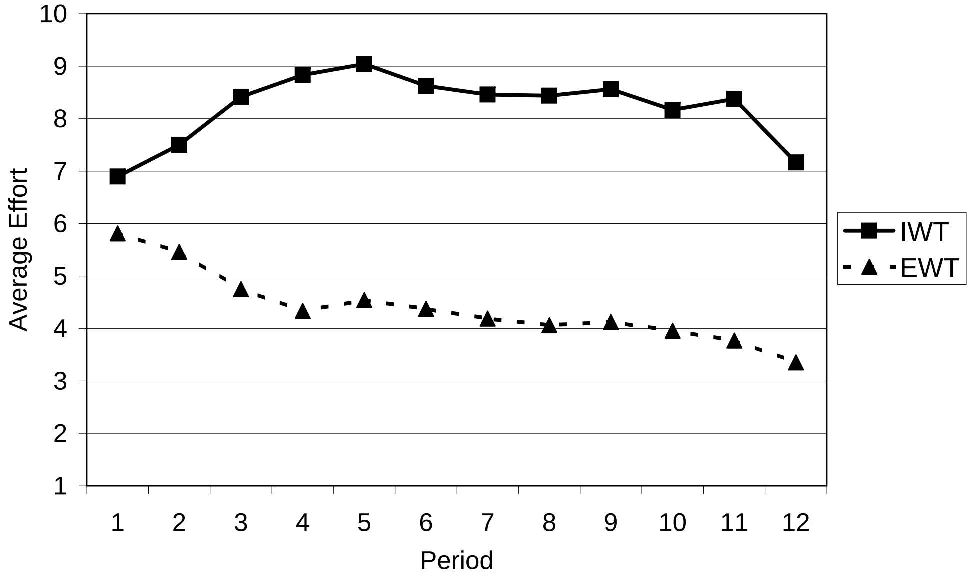


Figure 1.1: *Average effort per period. The effort is aggregated per period over all matching groups.*

$p < 0.01$). This means that the effort difference under the two wage schemes becomes even larger during the experiment. The treatment difference is also present when individual matching groups are considered: the highest average effort of an EWT matching group (5.88) is still lower than the lowest average effort of an IWT matching group (7.47).

The difference in agents' behavior can also be seen in the histogram of effort choices (Figure 1.2). In the individual wage treatment agents choose the maximum effort of 10 in 49% of the cases, 84% of the choices are higher than 6. Under equal wages, agents choose an effort higher than 6 in only 26% of all cases. The effort decisions are more spread out in the EWT, the minimal effort of 1 being the modal choice with 24% of the choices.

The comparison of effort levels across treatments shows that the enforcement power of reciprocity strongly depends on the wage scheme that is used. Under equal wages, efforts are relatively low, reaching only about half the level of efforts in the IWT. At the same time, the individual wage institution is very successful in eliciting efforts. Although contracts are not enforceable at all, efforts are close to the maximum in the IWT. Since higher efforts increase production and since the

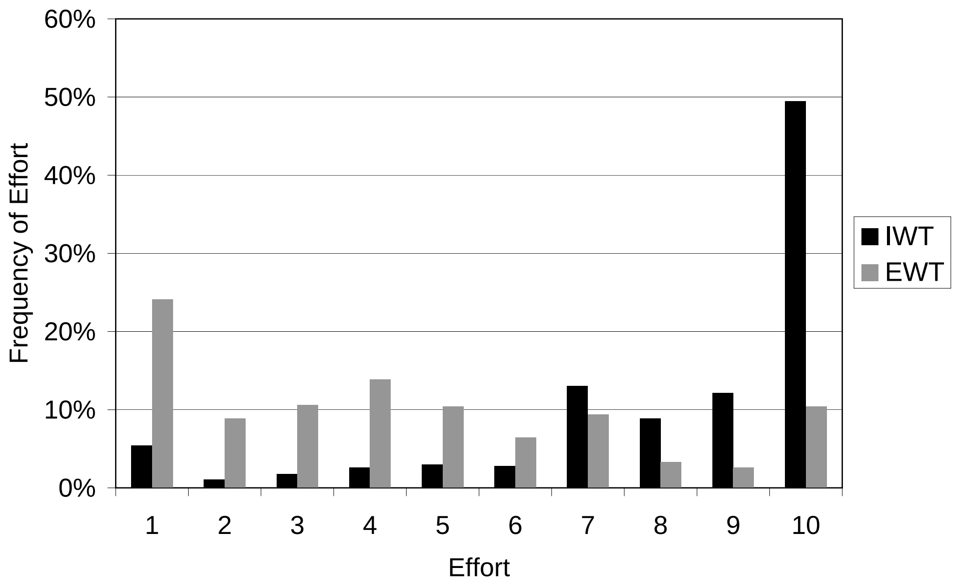


Figure 1.2: *Frequency of effort choices.*

marginal product of effort always exceeds its marginal cost, the differences in effort provision directly translate into differences in efficiency.

> **Result 1:** *The two payment modes exhibit strong differences with respect to the performance they elicit: agents who are paid equal wages exert significantly lower efforts than agents who are paid individually. This results in a much higher efficiency under individual wages.*

Both, the agents and the principals benefit from the increase in efficiency. The average period profit of a principal is 56 in the EWT compared to 100 in the IWT (Mann-Whitney test: $p < 0.01$), while agents earn an average period profit of 10 under equal wages vs. 17 under individual wages (Mann-Whitney test: $p < 0.01$).[10]

1.3.2 Wage Setting and Monetary Incentives

In order to better understand the vast differences in effort choices, we now take a closer look at principals' wage setting and the resulting monetary incentives for

[10]The large payoff difference between principals and agents is (at least partly) driven by to the two-to-one matching and the last-mover advantage of principals that has also been observed in other gift-exchange experiments (e.g., Fehr et al. forthcoming).

the agents under the two payment schemes. Figure 1.3 plots the average wage per effort level in the two treatments. For both treatments we take the wage paid by the principal for each individual effort decision and calculate averages for a given effort level.[11] The graph exhibits the upward sloping effort-wage relation of many gift-exchange experiments. For example, an agent in the equal wage treatment who exerts an effort of 1 receives on average a wage of 6.3 while an agent exerting an effort of 10 receives an average wage of 30.3. In the individual wage treatment, the corresponding wages are 1.7 and 39.5.

The effort-wage relation indicates that principals are indeed reciprocal, i.e., they reward higher effort levels with higher wages. While this holds true in both treatments, some differences between the treatments are worth noting. First, the wage increase is somewhat steeper in the IWT. Moreover, the average wages in the EWT do not rise as steadily as in the IWT but fluctuate more strongly, especially for high effort levels. While this may partly be due to the low number of high effort observations in the EWT, it might also be caused by a stronger influence of the co-worker's effort e_j on worker i's wage. We will turn to this point in more detail below.

> **Result 2:** *Principals reward a higher effort with a higher wage in both treatments.*

The reciprocal behavior of the principals generates monetary incentives for agents. The potential of reciprocity to enforce incomplete contracts partly depends on these monetary incentives. Therefore, we will now explore the impact of the purely monetary incentives on agents' behavior, while bearing in mind that also non-monetary aspects of the payment schemes will be important for the agents (see Section 1.3.3).

In order to derive the monetary incentives entailed in the principals' wage decisions, one first has to take into account that the agents have to pay the cost of

[11] Thus every wage decision of the principal enters twice in the equal wage treatment. In the IWT, principals can set two wages, and each of these wages enters the analysis once. Principals in the IWT do indeed use the possibility to set different wages. If efforts differ they also pay different wages in 91.4% of cases (see also Section 1.3.3).

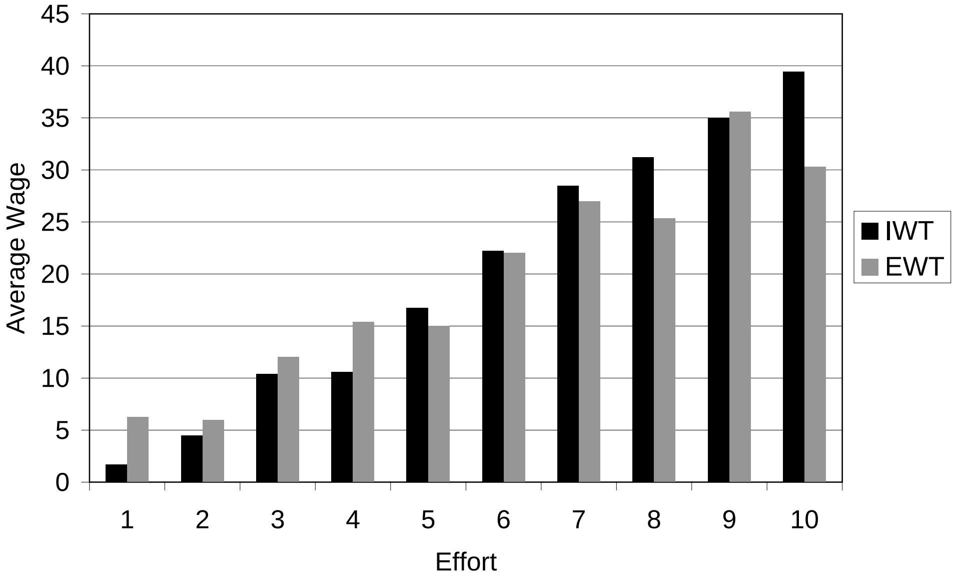


Figure 1.3: *Average wage for a given effort.*

effort exertion (see Table 1.1). Qualitatively, this does not change the picture of the effort-wage relation: higher effort levels seem to lead not only to higher wages, but also to higher profits for the agents. In order to check this in more detail, we estimate a simple linear OLS-model where we regress the agent's (period) profit π_{A_i} on his effort level e_i and a constant. To account for potential differences between the treatments we include a treatment dummy IWT, and an interaction term of the treatment dummy and agent's effort. IWT is equal to 1 for the individual wage treatment and equal to 0 for the equal wage treatment. The estimation results are reported in column 1 of Table 1.3.[12] The coefficients indicate that the effort-profit relation is indeed positive in both treatments. On average, an additional unit of effort increases the agent's profit under equal wages by 1.031 points. This coefficient is weakly significant. In the individual wage treatment the effort-profit relation is slightly steeper: an effort increase of 1 leads to an increase in agent's profit of 1.804 points (1.031 + 0.773). The difference between treatments, however, is not significant.

[12]We allow for dependent observations within matching groups and assume that only observations in different matching groups are independent. The reported robust standard errors are adjusted for this clustering.

We speculated above that exerting high effort levels might be more risky for an agent under equal wages since the co-worker's effort has probably a stronger influence on the principal's wage payment. An agent under individual pay only bears the risks of contractual incompleteness, i.e., he risks meeting a principal who is not reciprocal and pays no (or a low) wage. Under equal wages, the agent additionally faces the risk of receiving a lower wage because his co-worker negatively influences the principal's wage decision. This might weaken the incentives to provide high efforts, especially for risk-averse agents. To check whether high effort provision nevertheless pays off individually, we estimate a second model where we control for the co-worker's effort e_j (see Column 2 of Table 1.3).[13] The results indicate that indeed the co-worker's effort choice has a substantial influence on an agent's profit under wage equality while it has a negligible influence if individual wages are paid. An increase in agent j's effort increases agent i's profit in a given period by 2.774 points in the EWT, while the influence in the IWT is -0.404 $(= 2.774 - 3.178)$.[14] However, it is still individually profitable for the agents to exert high efforts in the EWT. An additional unit of (own) effort increases the agent's profit by 0.854 points.

The regression analysis suggests that exerting higher efforts is profitable for the agents under both wage schemes, at least if one averages over all observations in our sample. We now analyze whether this also holds if we aggregate efforts and profits individually for each subject. In the scatter plot shown in Figure 1.4, the x-axis depicts the average effort of an agent that he exerted over the course of the 12 periods, the y-axis shows the sum of all profits of an agent. Each dot thus represents one subject. The picture confirms the previous impression: subjects who exerted a higher average effort level during the experiment earned higher profits in total. More importantly, for the (few) observations where agents provide similar average effort levels in the two treatments earnings are very similar, too. In light of this, the

[13]In order to estimate the influence of the co-worker's effort e_j we have to split the sample such that only one observation per firm is included in the analysis. In order to make the two specifications comparable, we reported the first regression for the same sample. The results do not depend on which worker's effort is selected as "e_i".

[14]A separate regression for the IWT (not reported here) indicates that this value is not significant ($p = 0.219$).

Dep. Variable	π_{A_i}	π_{A_i}
e_i	1.031*	0.854**
	(0.535)	(0.348)
$IWT \times e_i$	0.773	0.995*
	(0.615)	(0.469)
$cons$	5.927**	-5.815***
	(2.614)	(1.523)
IWT	-3.744	11.004***
	(3.235)	(3.274)
e_j		2.774***
		(0.280)
$IWT \times e_j$		-3.178***
		(0.403)
N. Obs.	576	576
R^2	0.100	0.238

Table 1.3: *Profit regressions. Robust standard errors are given in parentheses. The dummy "IWT" is equal to 1 for the individual wage treatment. Significance at the 10%, 5% and 1% level is denoted by *, **, and ***, respectively.*

strong differences in actual efforts and especially the low effort levels under equal wages are remarkable.

One could object that the subjects in the experiment did not have access to the analyses we just presented. Both the regression analysis and the effort-profit relation of Figure 1.4 are "ex-post" examinations while the subjects only observed the behavior and outcomes of their previous groups. It could thus be that subjects were not able to learn that high efforts are profitable given the limited information they had. To explore if this is the case we calculate the profit-maximizing effort level for each agent in each period based on the information this subject actually has. The agent is assumed to choose the effort level that was on average the most profitable of all effort levels he has observed so far.[15] The calculation shows that agents in the EWT could have increased their efforts and profits considerably even by using only

[15]Since we assume that subjects do not "try" a never-observed effort level and since some subjects in the EWT never observe high effort levels in their group, this calculation *underestimates* the optimal effort level for the EWT.

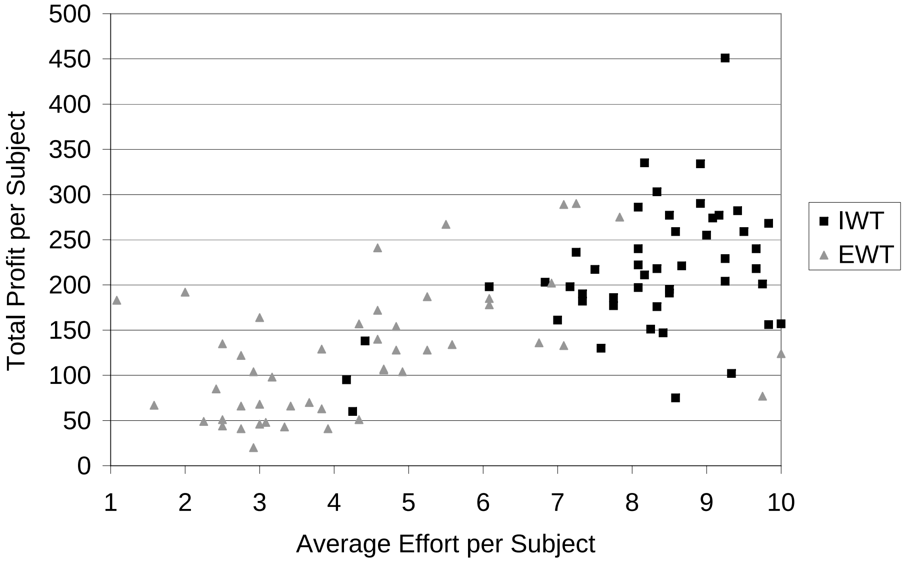


Figure 1.4: *Total profits of agents given their average effort level over all periods.*

their limited information. In the last period, the average profit-maximizing effort level exceeds the average actual level in that period by 61%. By contrast, subjects in the IWT do find the profit-maximizing effort levels: the average actual effort levels in the IWT are very close to the profit-maximizing levels. Our findings concerning agents' monetary incentives can be summarized as follows.

> **Result 3:** *The wages paid by principals imply similar monetary incentives in both treatments. A higher effort leads to a higher profit in both treatments. The increase is only slightly stronger under individual wages.*

While the analysis of the monetary incentives yielded some differences between the two wage schemes at hand, these differences can hardly explain the discrepancy in agents' performance reported above. Agents under individual wages provide very high effort levels, which is in line with the monetary incentives. On the other hand, agents under equal wages predominantly choose low efforts, thereby foregoing considerable profits. In light of these results, it is all the more important to analyze the non-monetary incentives of the two wage schemes in detail.

1.3.3 Non-Monetary Incentives

In the preceding section we presented evidence that many principals reciprocate a higher effort with a higher wage in both treatments. As discussed in Section 1.2.2 this is in line with the norm of reciprocity. However, reciprocal agents will additionally care about whether the worker who works more than his co-worker also receives a higher payoff than his colleague. A violation of this second aspect of reciprocity has different implications for the two agents involved. First, an agent who works more but does not receive a higher payoff than his co-worker suffers twice: he feels unfairly treated *and* he earns less. Thus, we refer to this situation as *disadvantageous norm violation*. Analogously, his co-worker who exerts a lower effort and earns a higher profit faces an *advantageous norm violation* since the unfairness is at least to his monetary advantage.[16] These norm violations cause non-monetary incentives that reinforce or counteract the monetary incentives implied by the wage setting. The combination of these two types of incentives will determine how agents perform under the respective wage scheme.

Agents' reactions to norm violations

In the following, we analyze how agents change their effort provision after experiencing an advantageous or a disadvantageous norm violation. We will show that a norm violation leads to an overall decrease in effort. This effect is very similar in both treatments, however, the norm of reciprocity is violated much more frequently in the EWT.

Table 1.4 shows how often agents decrease, increase or do not change their effort from period t to $t+1$ after they experienced no, an advantageous or a disadvantageous norm violation in period t. The top panel of Table 1.4 reports data for the equal wage treatment. When the norm is fulfilled, most agents keep their effort constant (54%) and slightly more agents increase their effort than decrease it. After

[16] More precisely, an advantageous norm violation comprises all cases when efforts are equal but payoff is higher, or when effort is lower effort but payoff is not. A disadvantageous norm violation occurs if efforts are equal but profit is lower, or if effort is higher but profit is not.

	Effort Down	Effort Constant	Effort Up	N. Obs.
EWT				
No Violation	19.1 %	54.4 %	26.5 %	68
Adv. Violation	12.2 %	43.5 %	44.3 %	230
Disadv. Violation	52.6 %	33.9 %	13.5 %	230
Total	30.7 %	40.7 %	28.6 %	528
IWT				
No Violation	19.2 %	51.8 %	29.0 %	448
Adv. Violation	45.0 %	27.5 %	27.5 %	40
Disadv. Violation	35.0 %	57.5 %	7.5 %	40
Total	22.3 %	50.4 %	27.3 %	528

Table 1.4: *Frequency of effort reactions.*

experiencing an advantageous violation of reciprocity, agents tend to increase their effort (44%) and only few reduce it (12%). The opposite is true after a disadvantageous norm violation: the majority of agents decrease their effort (53%) and only few increase their effort in the following period (14%). These numbers suggest that agents dislike being exploited (disadvantageous norm violation) and dislike feeling guilty (advantageous norm violation). After a norm violation they change their effort provision in the direction that makes a violation less likely to occur in the next period. This is consistent with the predictions of reciprocity.[17]

Behavior in the individual wage treatment (bottom panel) is very similar to behavior in the EWT for the cases of no violation and disadvantageous violations. When the norm is not violated agents mostly keep their effort unchanged. After a disadvantageous norm violation efforts are decreased rather than increased, as in the EWT. The only difference between the treatments is observed when agents experience an advantageous norm violation: agents in the IWT tend to decrease their effort while the EWT agents tend to increase it in this case.

[17]Similar effects are observed by Thöni & Gächter (2005) in a related set-up. They allow agents to revise their effort decision after learning their co-workers' effort choice. In the revision stage, the majority of agents decreases the effort difference to their co-worker, i.e. agents with initially higher effort revise their decision downwards while agents with lower effort revise it upwards.

If behavior is so similar between treatments, how can a preference for reciprocity cause the treatment effect? The last column of Table 1.4 shows how often the three situations occur in the two treatments. In the EWT, the norm is violated in 87% of all cases (460 out of 528) since this happens whenever agents exert different efforts. By contrast, in the IWT reciprocity is violated only in 15% of the cases (80 out of 528). Thus, even if the behavior in a given situation is similar, agents in the EWT are far more often exposed to norm violations than agents in the IWT. This is not caused by the principals per se. It is rather the heterogeneity in efforts combined with the equal wage institution that forces principals to set wages that are not in line with reciprocity. Principals in the IWT seem to understand the mechanisms of reciprocity quite well and use the possibility to set different wages in a sophisticated way. If efforts differ they also pay different wages in 91.4% of the cases, the more hard-working agent getting the higher wage in 98.8% of these cases. Additionally they do not treat agents differently if they exert the same effort: if efforts are equal, principals also pay equal wages in 90.1% of the cases.[18]

> **Result 4:** *Agents mostly react to disadvantageous violations of reciprocity by reducing their effort and by increasing it after an advantageous norm violation. The norm of reciprocity is far more often violated in the equal wage treatment.*

So far we have seen that agents' reactions are largely in line with the hypothesized behavior of a reciprocal agent and that treatments differ with respect to the frequency of norm violations. Yet, this is not sufficient to explain the treatment effect, since a norm violation is always advantageous to the one agent and at the same

[18]We checked the robustness of the reaction patterns in several ways. For example, it could be possible that the results are driven by strong dynamics at the beginning of the experiment or by an end-game effect. The results stay however very similar if one restricts the analysis to the first or the second half of all periods. It could also be that agents react differently to norm violations if they are paid very high or low wages. However, performing the analysis only for agents receiving a wage out of the top or bottom quartile of the ex-post wage distribution does not alter the result. An implicit assumption of our analysis is that the gift-exchange relation is generally intact between principal and agent, i.e., that agents exert a non-minimal effort and that principals pay a positive wage. The results do not change if one restricts the analysis to these cases. Also if one defines gift exchange as requiring the agent's profit to be positive, i.e. $w_i > c(e_i)$ instead of $w_i > 0$, the results are very similar.

time disadvantageous to the other one. If both agents adjust their effort in a similar way but in opposite directions the adjustments will cancel out. Reciprocity can only explain the downward trend in effort provision in the EWT if the reaction to a disadvantageous norm violation is stronger than the reaction to an advantageous one.

We therefore analyze the magnitude of agents' reactions to norm violations. Figure 1.5 shows the average change in effort provision from period t to $t+1$ after an agent experienced no norm violation, a disadvantageous or an advantageous norm violation in period t. The width of the bars corresponds to the number of observations in the respective category (cf. last column of Table 1.4). After a disadvantageous norm violation, agents in the EWT react strongly. They decrease their effort by 1.30. Their co-worker, experiencing an advantageous norm violation, increases his effort but not as strong. He raises his effort by only 0.75.[19] The direction of effort change is in line with the frequencies presented in Table 1.4. This analysis indicates that agents suffer more from a disadvantageous norm violation than from an advantageous violation. Thus the combination of a disadvantageous and an advantageous norm violation translates into non-monetary incentives that lead to an overall decrease in efforts.

As already observed above, in the IWT both groups of agents experiencing a norm violation decrease their effort. When reciprocity is not violated agents tend to keep their effort constant or even slightly increase it. We performed the same robustness checks as for the analysis of Table 1.4. All the alternative specifications yield results similar to the baseline specification depicted in Figure 1.5.

> **Result 5:** *Agents' reactions to a violation of reciprocity are asymmetric: the negative reaction of the disadvantaged agents is stronger than the positive reaction of the advantaged agents. This asymmetry in agents' reactions results in an overall negative time trend in efforts for the EWT and in the strong treatment difference in effort.*

The equal wage treatment leads to frequent norm violations. Agents experience

[19]The difference is statistically significant (Wilcoxon test of the absolute values: $p = 0.01$).

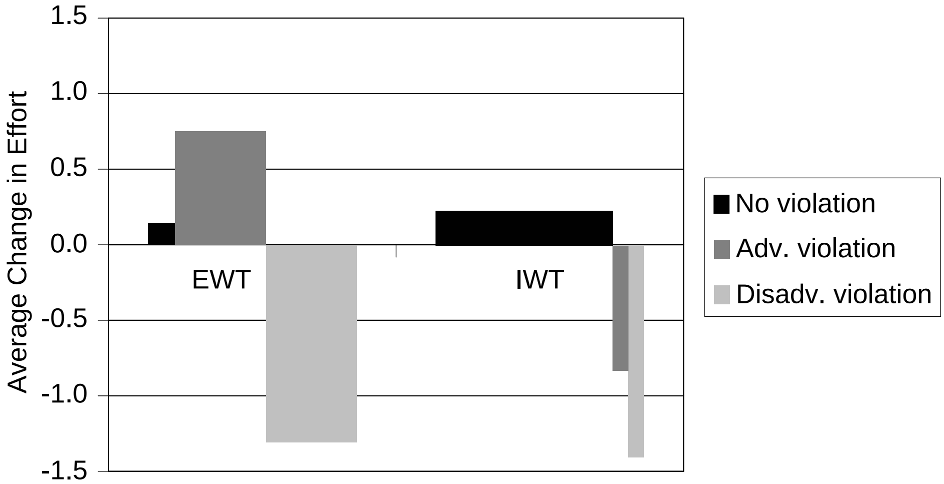


Figure 1.5: *Magnitude of effort reactions. The average change in effort from period t to period $t+1$ is shown given that the agent experienced no norm violation, an advantageous violation or a disadvantageous norm violation in period t. The width of the bars corresponds to the number of observations.*

the equal wage scheme as less fair.[20] Interestingly, even the principals consider the equal wage scheme as less fair. In the post-experimental questionnaire, principals are presented three hypothetical game situations that include effort choices, wage choices, and the resulting payoffs for all players. They are asked whether they consider the resulting allocation as just. One of the three situations reflects their own average behavior in the experiment.[21] The principals do not know that they are facing their own past decisions when answering this question. 63% of the principals in the IWT consider their own decisions fair while only 38% of the principals in the EWT share this view (Mann-Whitney test on matching group shares: $p = 0.03$).

[20]Note that the treatment effect cannot be explained by profit inequalities per se. The absolute differences between co-workers' payoffs are not significantly different between treatments (IWT: 6.47, EWT: 7.14, Mann-Whitney test: $p = 0.29$) but the sign differs: in the IWT, the harder working agent earns more, while the opposite is true in the EWT. Apparently, different profits are not considered as unfair as long as the hard-working agent gets the higher payoff.

[21]This situation was constructed as follows: We calculated the average effort of the higher-effort and of the lower-effort providers that the principals actually faced during the experiment. We then took the average of the wages the principals paid to the two groups. Finally, we calculated hypothetical payoffs for all three "average" players by considering the costs of the average efforts.

Summarizing, one can say that non-monetary incentives differ between the treatments. The equal wage institution forces the principals to violate reciprocity every time efforts are different. These norm violations translate into non-monetary incentives, partly overpowering monetary incentives and causing adverse reactions by the agents. The asymmetry in the strength of reactions to a norm violation, especially the strong negative reaction to a disadvantageous one, is then able to explain the overall negative effort trend and ultimately the low effort levels in the EWT. By contrast, in the individual wage treatment agents perform well since they are content with the fair treatment: the more hard-working agent earns almost always more than his co-worker. Principals use reciprocity forcefully as an incentive device, inducing high performance of agents. Thus, all parties gain in monetary and non-monetary terms. The individual wage scheme is not only more profitable but also experienced as fairer.

Simulation with reciprocal agents

We have seen so far that the presence of reciprocal agents in combination with the frequent violations of reciprocity in the EWT are able to explain the treatment difference. In order to further illustrate how institutions and reciprocity interact, we take our previous findings on agents' period-to-period reactions and link them to the aggregate dynamics in the experiment. We do so by simulating agents' behavior with a simple "reciprocity adjustment" rule. In this simulation, all agents are assumed to derive utility from money, but to also suffer from violations of a norm of reciprocity. When deciding about their effort in a given period, the simulated agents therefore compare their effort and profit in the previous period with the effort and profit of their co-worker in that period. According to the comparison along these two dimensions, four reactions can be distinguished for the simulated agents. (i) For an agent who had a higher effort and a higher profit, reciprocity is not violated and the pecuniary comparison is also advantageous for him, so he keeps his effort constant. (ii) For an agent who exerted a lower effort and got a lower profit, the norm of reciprocity is satisfied but profit maximization is not, thus he partly adjusts his effort in the direction of his co-worker's effort, i.e., he chooses an effort $(e_{i,t} + e_{j,t})/2$.

(iii) An agent with higher effort and lower profit feels exploited as he suffers from a disadvantageous norm violation. Thus he adjusts his effort fully and chooses $e_{j,t}$. (iv) Finally, for an agent with lower effort and higher profit the norm violation is advantageous, thus the resulting utility is higher than in case (iii). He chooses an effort $(e_{i,t} + e_{j,t})/2$. The reactions in cases (i) to (iv) are in line with the period-to-period reactions presented in Table 1.4 and Figure 1.5.

In the simulation, we use actual effort data from the experiment only for the first period. The subsequent effort decisions are based on the simulated profits and simulated efforts of the previous period. The simulated principals pay the average wage for a given effort (IWT) or the average wage sum for a given effort sum (EWT) as calculated from the experimental data. Profits are then calculated as wage minus cost of effort exertion. We use the same matching protocol as in the experiment.

Figure 1.6 shows how effort choices evolve over time in the experimental data and in the simulations. The simulations *'EWT sim'* and *'IWT sim'* trace the real data very well and are able to reproduce the large effort difference between treatments. In the individual wage simulation, efforts increase like the real efforts although the slight downward trend in the second half of the experiment cannot be reproduced. Efforts in the equal wage simulation constantly decrease down to an effort level slightly above 3 in the final period. This pattern is very similar to the dynamics in the real data.

We performed several robustness checks of the simulation. To check how the first period efforts influence the result, we initialized the EWT simulation with the first period efforts of the IWT agents and the IWT simulation with EWT-first-period-efforts. The dynamics are very similar to the baseline specification: in the IWT simulation, agents steadily increase their efforts while they decrease their efforts in the EWT simulation. Also other first-period-effort-vectors do not change the general result of the simulation as long as efforts are sufficiently heterogeneous. If agents exerted the same effort already in the first period, equal wages would not violate reciprocity and efforts would not decrease over time. The rules specified for cases (i) to (iv) do impact the result, though not all with the same strength. While the rules for case (ii) and (iv) can be changed without altering the qualitative result of the simulation very much, it is crucial to keep the rules for case (i) and (iii) similar to our baseline specification. The treatment difference can only be reproduced in this setup if the agents revise their effort decision downwards after experiencing a disadvantageous norm violation (case (iii)) and if they continue to exert high efforts when they earn money and feel treated fairly (case (i)).

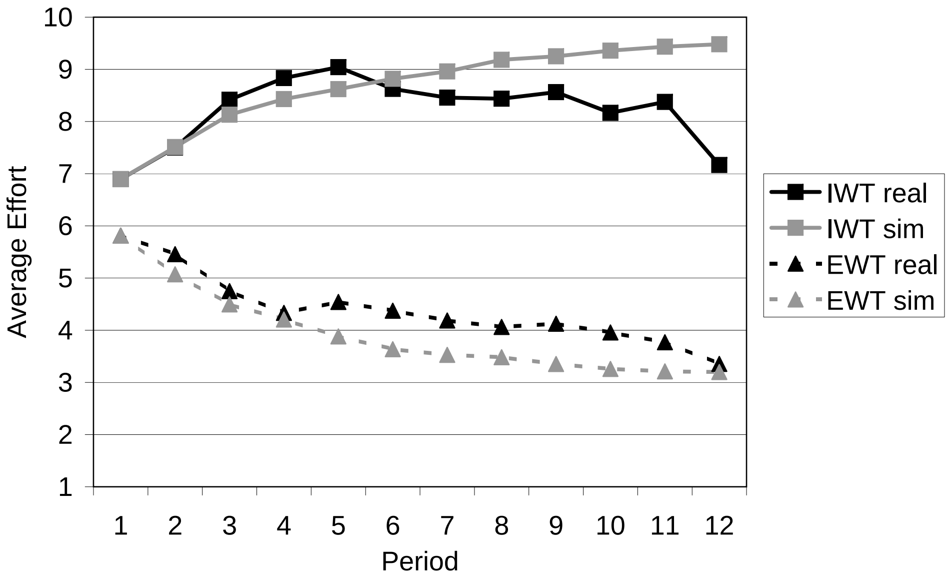


Figure 1.6: *Simulation with reciprocal agents.*

Note that the pivotal agent is different between the simulated treatments: in the equal wage simulation the norm of reciprocity is not fulfilled when agents choose different effort levels. In these cases in the simulation, the agent with the higher effort will fully adjust his effort in the direction of his co-worker's effort while the co-worker will increase his effort level only to the average effort of the last period. In the EWT simulation, the average effort therefore converges to the lowest first period effort as agents are subsequently matched together: the low-effort providers are pivotal. By contrast, in the IWT the high-effort providers have the decisive impact on the overall outcome. The norm of reciprocity is mostly fulfilled in the IWT. Thus, the agent with the higher effort keeps his effort constant while his co-worker adjusts his effort. The average effort therefore converges to the highest first period effort. We will analyze this point in more detail in the next section.

Result 6: *Simple simulations based on agents who have preferences for money and reciprocity are in line with the efforts observed in the experiment and are able to reproduce the observed treatment effect.*

1.3.4 Dynamics of high- and low-effort providers

As already seen in Figure 1.2, subjects exhibit a substantial degree of heterogeneity with respect to effort provision. In the following, we analyze if the agents who are most or least willing to exert effort are affected differently by the two payment modes at hand. A common informal argument claims that equal wages will be especially detrimental to the motivation of high performers (see e.g., Milgrom & Roberts (1992), p. 419) but clean empirical evidence is scarce. Furthermore, it is unclear how weakly motivated agents react to equal or individual wages. We also address the question whether high and low performers impact the overall results differently in the two treatments. The simulations presented in the previous section suggest that this could indeed be the case: in the EWT simulation, the low-effort providers are decisive for the final outcome while it is the high-effort providers in the IWT simulation.

To analyze these questions in the experimental data we classify agents according to their effort decision in the *first period*. We define the agent with the highest first-period effort in each matching group as "high-effort provider" and the agent with the lowest effort as "low-effort provider".[22] This type definition is chosen because when agents decide on their effort in the first period, they do not have any information about the behavior of other subjects and all learning and coordination processes occur after this initial effort choice. Thus first-period effort is likely to be a good proxy for the intrinsic willingness of a specific agent to exert effort. If some of the subjects are intrinsically inclined to exert high efforts they should show up in the group of high-effort providers. In contrast, if some of the subjects are intrinsically inclined to exert low efforts they should show up in the group of low-effort providers.

In Figure 1.7 we follow the high-effort providers and low-effort providers in both

[22] If more than one agent chooses the highest or lowest effort in the first period, the subsequent effort decisions of these agents are averaged.

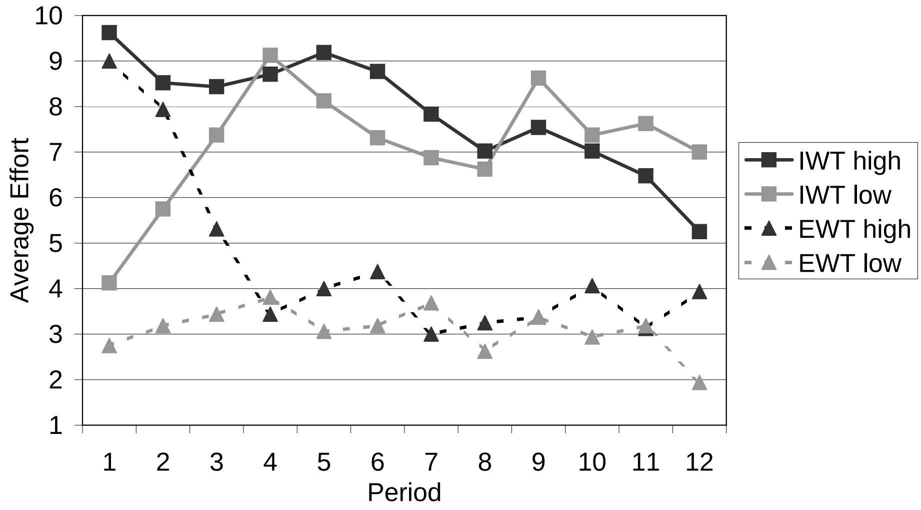


Figure 1.7: *Effort decisions of high-effort and low-effort providers. In each matching group, the agent with the highest (lowest) effort in the first period is defined as the high (low)-effort provider.*

treatments and show their effort decisions over time. In the first period, the groups of high-effort providers and the groups of low-effort providers are close together across treatments.[23] This changes completely over the course of the 12 periods. In the individual wage treatment, high-effort providers continue to provide high effort levels. Low-effort providers increase their efforts dramatically up to the level of the high-effort providers and even higher in the last periods. In the equal wage treatment, the dynamics are reversed. Here, the low-effort providers keep their effort provision constant and the high-effort providers reduce their efforts to the level of the low-effort providers.[24] Put simply, the "good" agents push the "bad" agents up under individual wages while under equal wages the "bad" ones pull the "good" ones down.

[23]In the first period, effort levels are not significantly different between treatments for high-effort providers (Mann-Whitney test: $p = 0.14$) while they are close together but different for the low-effort providers (Mann-Whitney test: $p = 0.03$). Within treatments, the high-effort and low-effort providers choose statistically different effort levels in the first period (Wilcoxon signed rank test: $p = 0.01$ (IWT), $p = 0.01$ (EWT)).

[24]In the last six periods, effort levels are not different within treatments (Wilcoxon signed rank test: $p = 0.67$ (IWT), $p = 0.78$ (EWT)) while they differ between treatments (Mann-Whitney test: $p < 0.01$ (high-effort providers), $p < 0.01$ (low-effort providers)).

These dynamics underline the importance of the different non-monetary incentives induced by the two wage setting institutions. Remember that agents face similar monetary incentives in both treatments. Wage equality often violates reciprocity. Agents in this treatment who are in principle willing to exert high levels of effort get frustrated and lower their efforts. On the contrary, under individual wages where reciprocity is intact, good performance spreads. These results suggest that choosing a wage scheme also influences the social dynamics between the agents.[25] In our experiment, individual wages lead to positive dynamics since agents orient themselves to the most hard-working agents. In contrast, the equal wage scheme focuses agents' attention on the least motivated agents.

> **Result 7:** *The pivotal agent is different between treatments: in the IWT the initially low-effort providers align with the high-effort providers over time. In the EWT the initially high-effort providers align with the low-effort providers over time.*

1.4 Conclusions

In this chapter, we studied the interaction of reciprocity with different wage schemes. More specifically, we analyzed how effective equal wages are in an environment where contract enforcement solely relies on reciprocity. In our experiment, one principal is matched with two agents. The principal pays equal wages in one treatment and can set individual wages in the other. The use of equal wages elicits substantially lower efforts and efficiency in spite of similar monetary incentives for the workers under both wage schemes. In particular, exerting high effort pays off in both settings. The strong treatment difference is driven by subjects' reciprocal preferences and the fact that reciprocity is frequently violated in the equal wage treatment. This is not the case in the individual wage treatment, as principals set wages mostly in line with reciprocity.

Our results have a number of implications, both for the advancement of existing theories and for the design of wage schemes in practice. First of all, it is doubtful

[25] See Manski (2000) for a discussion of social interaction effects.

whether wage equality can be reconciled with the use of reciprocity to enforce incomplete contracts. Our findings rather suggest that the possibility to individually sanction bad performance and reward good performance is a crucial prerequisite to make reciprocity the powerful enforcement device it has proven to be in many bilateral interactions. The performance of agents in the individual wage treatment shows how effective reciprocity can be: although explicit contract enforcement is absent, 80% of the possible efficiency gains are realized.

Second, while it is well-known that equal wages can distort monetary incentives, in our experiment they are also perceived as less fair and thus efficiency decreasing, even though the monetary incentives are qualitatively not affected. This holds in particular because agents differ in their performance. It may thus be oversimplifying to argue that equal wages lead to less envy and therefore higher work morale, as it is frequently done in the political discussion.

Third, in practice the discretion to individually reciprocate good performance of subordinates does not have to be in monetary terms. Non-monetary benefits like extra vacation or awards can be useful devices to motivate workers in this context. These instruments become especially important when it is not possible to wage discriminate on a given hierarchical level, e.g., because the firm's internal pay structure, agreements with a union or legislation dictate wage equality.

The results in this chapter should not be interpreted as arguments against wage equality in general. They rather suggest that equal wages come at a cost that has to be weighed against their potential benefits. For example, equal wages are easier to implement than individual wages, and they may encourage peer monitoring and collaboration. The relative importance of these costs and benefits (and also the impact of the workforce's social preferences more generally) is likely to depend on the details of the institutional setting. These include the production technology, the information structure, and the organizational design of the firm. In this chapter we presented results for one such setting. Our design provides a simple and parsimonious framework that can successively be enriched to study these aspects in future research.

Chapter 2

Fungibility, Labels, and Consumption

2.1 Introduction

Fungibility of money is a central principle in economics. It implies that any unit of money is substitutable for another. In the analysis of consumer choice, for example, fungibility prescribes that consumption decisions are based exclusively on the consumer's total wealth—its composition is irrelevant (Modigliani & Brumberg 1954). Fungibility is assumed throughout most of economic theory. Some empirical findings, however, cast doubt on the generality of the concept. Odean (1998), for example, analyzes stock market behavior of individual investors and finds that investors sell winning stocks too soon and keep losing stocks too long. He explains this finding by assuming that investors evaluate each stock holding separately (i.e., treat them as non-fungible) and are loss averse in each stock holding with respect to the buying price. Other examples come from the fields of asset pricing (Benartzi & Thaler 1995, Barberis et al. 2001), stock market participation (Barberis et al. 2006), stock trading (Barberis & Huang 2001), and life-cycle saving (Shefrin & Thaler 1988). Despite the fundamental importance of fungibility, surprisingly little is known about the degree to which individual decision-making is in line with the notion of fungibility.

In this chapter, we investigate whether individuals treat different income sources as fungible using a combination of a natural field experiment and a controlled laboratory experiment. Both studies have the same general design: the consumer has a cash budget and an additional lump-sum subsidy of amount S at his disposal. In the *Cash treatment*, the subsidy S is given in cash. In the *Label treatment*, the subsidy is given as in-kind benefit, i.e., the subsidy has to be spent on the subsidized good. The crucial feature of our design is that the optimal consumption of the subsidized good exceeds the amount of the subsidy. By shifting the remaining budget, the consumer can reach the same first-best consumption level in both treatments, i.e. the subsidy is inframarginal and not distortionary. Therefore, treatments merely differ in the label attached to the subsidy; rational consumers should not be influenced by whether the subsidy is given as cash or in kind. Standard theory thus predicts consumption to be the same in both treatments. In contrast, a consumer who does not treat different income sources as fungible will spend the in-kind subsidy fully on the subsidized good and thus consume too much of this good. If consumption of the subsidized good is higher in the Label treatment, we can conclude that behavior is not in line with the principle of fungibility.

The field experiment was conducted in a wine restaurant. Guests received either a voucher for the whole bill (Cash treatment) or a voucher of the same amount that had to be used for beverage consumption (Label treatment). The value of the voucher was lower than the usual beverage consumption of almost all guests. In a restaurant, guests consume at least a minimal amount of two distinct goods (eating and drinking). Thus, an in-kind voucher of a lower amount than this minimal amount does not distort the optimal decision. The field experiment allows us to collect natural decision data in a controlled way. In the terminology of Harrison & List (2004) the experiment is a "natural field experiment": participants acted in a natural, incentivized, and well-known environment. Additionally, they were not aware that they participated in an experiment and therefore did not feel observed. They could not self-select into treatments as these were exogenously assigned. Participants could not even self-select into the experiment in general since vouchers were not advertised and came as a surprise to participants after they had entered the restaurant.[1]

The laboratory experiment has the same general design as the field experiment and offers an even more controlled and well-defined setup. Subjects could consume two goods and had at their disposal a cash budget and either a cash subsidy (Cash treatment) or an in-kind subsidy (Label treatment).[2] We induced a standard microeconomic utility function by specifying monetary payoffs for the possible consumption bundles. Importantly, the parameters were chosen such that the in-kind subsidy is inframarginal: the amount of the subsidy is lower than the amount spent on the subsidized good in the (first-best) optimum for every subject. This is a major advantage compared to field data where the exact utility function cannot be known and the individual budget varies. Since we know the optimal decision in the laboratory, we can also compare actual decisions to the optimum and calculate an individual error. Additionally, it is possible to collect further information about subjects and to analyze the channels of a potential treatment effect. Field and laboratory experiment are thus methodological complements.

The results of both experiments show that fungibility is violated in the settings under investigation. In the field experiment, consumption of the subsidized good (beverages) is considerably higher in the Label treatment than in the Cash treatment. Guests in the Label treatment spend 3.90 euro or 25% more on the subsidized good than guests in the Cash treatment. This effect is also large compared to the value of the voucher of 8 euro. The treatment difference is due to the choice of more expensive beverages: the average price per liter of consumed beverages is 3.52 euro higher in the Label treatment compared to the Cash treatment. In contrast, the treatment difference in consumed beverage *volume* is negligible. The total consumption (including meals) is not affected by the type of subsidy. But if we

[1]Self-selection because of word-of-mouth (i.e., early participants telling their neighbors) is very unlikely: we find that behavior is not different after the experiment compared to before. Since guests did not know when the experiment ended, a self-selection induced effect should show up also after the experiment.

[2]To have an additional intra-person measure of fungibility, subjects decided twice in the laboratory experiment. In a first stage, they allocated only the cash budget without the subsidy. Then the main treatment stage followed.

compare guests in the Cash and in the Label treatment with guests who did not receive either voucher, we find that spending is higher by 3.93 euro per person when a voucher is given.

The laboratory experiment confirms and extends the results of the field experiment. While average consumption of both goods is close to optimal in the Cash treatment, consumption of the subsidized good is significantly higher in the Label treatment. Subjects in the Label treatment consume on average 15% more of the subsidized good. Compared to a baseline stage without subsidy, subjects in the Label treatment spend 57% of the subsidy on the subsidized good compared to only 28% in the Cash treatment. One in five subjects in the Label treatment even spends every additional money unit on the subsidized good, while in the Cash treatment only one in fifty subjects does so. As we are able to collect more data on subjects in the lab than in the field experiment, we can analyze the underlying reasons for the treatment effect. We find a strong impact of subjects' mathematical abilities: subjects with higher mathematical skills (measured by their high school math grade) act consistently with standard economic theory. By contrast, in the group of subjects with lower mathematical skills, the treatment difference is large and significant. This supports the view that the violation of fungibility occurs for cognitive reasons and relates our study to recent work by Frederick (2005), Benjamin et al. (2006), and Casari et al. (2007). These studies show that people with higher mathematical and cognitive skills are more likely to behave in line with standard economic theory, whereas people with lower cognitive skills tend to act in accordance with theories of boundedly rational behavior.

Taken together, our results show that consumers do not always treat money as fungible. This has implications for several areas of economic research. First, our findings lend support to field studies that explain behavior of stock market investors by assuming that investors are loss averse in each portfolio position (e.g., Odean 1998, Benartzi & Thaler 1995). These studies assume that investors evaluate each component of their portfolio (e.g., stock vs. bond holdings) separately. We provide a direct test of this assumption. Second, our results are important for theories of life-cycle saving which usually rely on the assumption of fungibility.

Since we find that even in a very simple setup fungibility does not hold for all subjects, our evidence supports savings models in which the assumption of fungibility is relaxed (see, e.g., Shefrin & Thaler 1988, Barberis & Huang 2001). Furthermore, the specific design of our study allows us to give a rationale for the observed behavior of benefits recipients. Assuming a lack of fungibility could explain why housing benefits have such a strong effect on market rents as shown by Susin (2002), Gibbons & Manning (2006), and Fack (2006). Taking our results at face value, tenants receiving inframarginal housing benefits should be willing to pay a higher rent for a given apartment compared to tenants who get the same amount as cash grant. If landlords are aware of this behavior, they can exploit the existence of housing benefits to increase the rent. We discuss this issue in more detail in Section 3.5.

But why should people treat money as non-fungible? Tversky & Kahneman (1981) suggest that decision makers often do not decide globally but rather evaluate parts of a decision separately. This phenomenon has been called "narrow framing" (Kahneman & Lovallo 1993) or "narrow bracketing" (Read et al. 1999*b*). For the allocation of a budget coming from different sources, making separate decisions implies a violation of fungibility. Since assessing the decisions separately is cognitively less demanding, our finding that subjects with lower mathematical skills are more likely to violate fungibility also points to narrow bracketing as a potential explanation for the treatment effect. More directly related to the consumption setup we analyze is the concept of "mental accounting" (Thaler 1980, 1985, 1999). Mental accounting proposes that consumers use a set of heuristics to deal with their day-to-day financial decisions. Mental accounting assumes that consumers have mental budgets for different expenditure categories or for different investment categories, thereby constraining the fungibility of money. In this framework, a label can influence consumption choice if it determines to which mental budget the consumer assigns the benefit payment.

Most empirical studies testing fungibility rely on non-incentivized surveys (e.g., Heath & Soll 1996, O'Curry 1997, Prelec & Loewenstein 1998, White 2006). Only few papers investigate fungibility in incentivized laboratory experiments or with field data. Levin (1998) finds in a large household survey that the marginal propensity to consume out of current income is higher than out of wealth; he can exclude liquidity constraints as explanation. Milkman et al. (2007) show that customers of an online grocer spend more when they redeem a $10-off coupon. $10 is negligibly small compared to life-time wealth and should not alter spending behavior if customers treat wealth and coupon as fungible. Arkes et al. (1994) find that people spend more when they just received a surprise gift compared to when they expected to get the gift (see also Epley et al. 2006). In these three studies, however, the treatment effect could potentially be influenced by a change in mood because of receiving a gift. In our study, this cannot have an impact. Finally, Gneezy & Potters (1997), Thaler et al. (1997), and Rabin & Weizsäcker (2007) show in experiments that subjects evaluate subsequent gambles separately, i.e., treat them as non-fungible. There are, however, also a couple of studies finding that subjects do treat different income sources as fungible: for example, in an experiment by Maciejovsky et al. (2001), subjects do not treat income differently that they receive from selling experimental assets or from the assets' dividends; and Moffitt (1989), studying the food stamps program in Puerto Rico, finds no difference in consumption between households who receive inframarginal in-kind benefits or cash grants.

The remainder of the chapter is organized as follows: The design of both experiments is described in Section 3.2. Section 2.3 reports results of the field experiment. Section 2.4 presents results of the laboratory experiment and analyzes potential channels of the effect. Section 3.5 discusses an application of our results to benefit payments and concludes.

2.2 Experimental Design

2.2.1 General Setup

The goal of this chapter is to experimentally test whether individual behavior is in line with fungibility. We examine this question in a simple two-goods consumption case where one good is subsidized in a particular way. Assume that a consumer has a cash budget of amount R and a subsidy of amount S at his disposal. In the

Cash treatment, the subsidy S is given lump-sum in cash. In the *Label treatment*, the subsidy has the same amount and is also given lump-sum but it has to be spent on the subsidized good. The crucial detail of our design is that the optimal consumption of the subsidized good is larger than the amount of the subsidy, i.e., a rational consumer should not be influenced by whether the subsidy is given as cash or in kind.

Consider the indifference curve diagram in Figure 2.1, where the subsidized good (s) is on the horizontal axis and the other good (o) is on the vertical axis. For simplicity, the price of the subsidized good is normalized to $p_s = 1$. The dashed line is the budget constraint in the Cash treatment. Assume that the optimal consumption bundle is B. In the Label treatment, the subsidy is paid in kind; the consumer faces a kinked budget constraint (solid line). However, the kink does not affect optimal decision making, as the amount of S is lower than the amount s^B spent optimally on the subsidized good. The first-best choice B is still feasible. Under the assumption that fungibility holds, consumption should therefore be identical across treatments.

Now consider a consumer who does not treat different income sources as fungible. A violation of fungibility implies that the consumer has some sort of cognitive or mental sub-budgets between which money cannot be shifted (Thaler 1985). We assume that a labeled payment is posted to the sub-budget the label corresponds to, whereas a cash budget is allocated optimally to the different sub-budgets. In the Cash treatment, we would still expect such a consumer to choose bundle B. As both income components are cash, the consumer can allocate the optimal amounts to the sub-budgets for the two goods and thus choose the optimal consumption bundle. The difference to the standard model occurs in the Label treatment. The consumer will still allocate the cash endowment optimally (assume that this corresponds to bundle A). The subsidy, however, will be allocated to the sub-budget for the subsidized good. Since the sub-budgets are non-fungible, the subsidy will thus be spent entirely on the subsidized good. In the extreme case of complete non-fungibility, this results in a consumption of bundle C where $s^C = s^A + S$ (see Figure 2.1). If both goods are normal, $s^C \geq s^B$.[3]

Therefore, if some subjects do not treat money as fungible, we should expect average consumption of the subsidized good in the Label treatment to be higher than in the Cash treatment. This does not exclude the possibility that some subjects act in line with fungibility or that others are only influenced to a certain extent by the label attached to the subsidy.

It could be that people consume differently when they receive a gift or subsidy, just because they are in a positive mood. Such a change in mood cannot influence our main treatment comparison as subjects in both treatments get a subsidy.

2.2.2 Design of the Field Experiment

The field experiment was conducted in a wine restaurant situated in a wine-growing region of southern Germany. The restaurant itself is located in a winery. We chose this restaurant because almost all guests eat something *and* drink wine and other beverages; thus guests consume at least a minimal amount of two distinct goods.[4] Giving a beverage voucher that is smaller than the minimal amount spent for beverages will therefore not distort the optimal decision. Usual per-capita spending in this restaurant is about 40 euro (∼54 USD).

Guests were not aware of participating in an experiment. Upon arrival at the restaurant, they learned that the restaurant was celebrating its fourth anniversary (which was indeed the case) and that they would receive an 8-euro voucher per

[3]This reasoning depends on the order in which cash budget and subsidy are spent. If the consumer spent the subsidy first, he would be able to allocate the cash budget so as to reach bundle B. In the experiments, we are therefore testing the joint hypothesis of fungibility and order of spending. To minimize the influence of the order of spending as much as possible, subjects in the lab experiment decided twice: first, they allocated only the cash budget R. We will call this stage *reference stage*. In a second step, called *subsidy stage*, subjects allocated the combined budget $R + S$. The experience from the reference stage will probably lead most subjects to first allocate the cash endowment during the subsidy stage and then the subsidy, as they have already calculated how to allocate the cash budget R. Additionally, this setup allows us to calculate the consumption increase from reference to subsidy stage to get an intra-person measure of fungibility.

[4]In Germany, it is very unusual to not consume beverages in a restaurant; water must also be purchased.

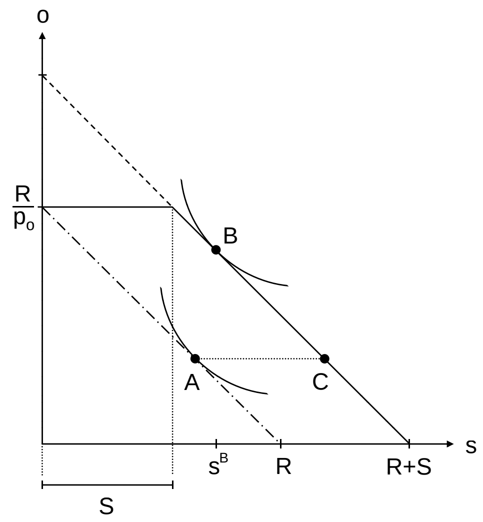


Figure 2.1: *Consumption decision with inframarginal in-kind benefit. The subsidized good (s) is on the horizontal axis, the other good (o) is on the vertical axis. The dashed line is the budget constraint when the subsidy is given in cash. The solid line is the budget constraint when the subsidy is given in kind. The dot-and-dash line is the budget constraint when no subsidy is given.*

person (∼11 USD at the time of the experiment). The type of voucher differed by day. On days of the Cash treatment, vouchers were given as "gourmet voucher" that could be spent on either beverages or the meal. To avoid confusion with the laboratory experiment, we will call this treatment *Field Cash treatment* (FCT). The Field Cash treatment serves as our primary control treatment. On days of the Label treatment, vouchers were given as a "gourmet beverage voucher". These vouchers could only be spent on beverages. We will refer to this treatment as *Field Label treatment.* From communication with the owner of the restaurant we knew that almost all guests consume beverages worth more than 8 euro, even without getting a voucher. Therefore, the beverage voucher should not distort the consumption decision of these guests; the only treatment difference should be the label on the voucher.[5] Both types of vouchers had to be redeemed the same evening.

[5]We could not avoid that some participants consumed less than the amount of the subsidy because total consumption had a high variance, ranging from 18.95 euro to 84.25 euro. In the laboratory experiment, described in the in next section, we can ensure that the subsidy is inframarginal for all subjects by choosing an appropriate payoff function and by endowing every subject with the same budget.

The restaurant distributed a total of 196 vouchers, one per person, starting with beverage vouchers until half of the available vouchers were distributed. To avoid distributing vouchers from both treatments on the same evening, the restaurant continued with beverage vouchers during that evening. From the next day on, the remaining vouchers were issued as bill vouchers. Between treatments, observations are thus counter-balanced over days of the week. Overall, 107 vouchers were distributed in the FLT and 89 vouchers in the FCT. We consider each table in the restaurant as one independent observation and calculate all absolute values per capita. Since we distributed one voucher per person, we can relate per-capita consumption directly to the amount of a single voucher. This leaves us with 37 independent observations in the Field Label treatment and 34 in the Field Cash treatment. No guest participated in both FLT and FCT. During the experiment, the menu did not change and the same two waiters were present in the restaurant. Our data consist of the detailed bill per table showing all consumed items and an additional questionnaire filled in by a waiter stating how many persons correspond to this bill, the share of women at the table, the share of persons below wine-drinking age, whether guests paid separately and the amount of the tip.

Additionally, we collected information about the time directly before and after the experiments to compare behavior in the two main treatments described above to the behavior of guests who did not receive either voucher. We have data on 356 persons which corresponds to 116 independent observations (76 before and 40 after the two main treatments).[6]

Results of the field experiment are presented in Section 2.3.

2.2.3 Design of the Laboratory Experiment

In the laboratory experiment, subjects had to make two subsequent consumption decisions. In each stage, subjects were endowed with a budget that they could

[6]We have the full information described above (bill and questionnaire) only for observations after the two main treatments. For the observations before, we know only the bill and the number of persons. Furthermore, it could be that some of these guests show up more than once in our data.

allocate on two goods. For each good, we induced a standard microeconomic utility function by specifying monetary payoffs for the possible consumption levels. Total payoff was the sum of the payoffs for each of the two goods in both stages. The first decision stage, which we will call *reference stage*, serves to yield a reference transaction to which decisions in the second stage can be compared. The second stage, called *subsidy stage*, is our main treatment stage.

In the reference stage, subjects received a cash budget of 50 money units which they could allocate freely on the two goods; the reference stage was identical in both treatments. In the subsidy stage, subjects again had an endowment of 50 money units at their disposal and additionally received a subsidy of 30 money units. The only difference between the two treatments was the type of the subsidy. In the *Lab Cash treatment* (LCT), the subsidy was given as an unconditional cash grant. In the *Lab Label treatment* (LLT), the subsidy was given as an in-kind benefit, i.e., the money had to be spent entirely on the subsidized good. Parameters were chosen such that the in-kind benefit was inframarginal for all subjects and not distortionary. By shifting the remainder of their budget appropriately, subjects could reach the same optimal consumption bundle as in the LCT. The only treatment difference therefore was the label attached to the subsidy.

The exact specification of the payoff functions is presented in Table 2.1. For each good, the payoff is increasing in consumption and the marginal payoff is weakly decreasing. Prices per unit were $p_s = 3$ for the subsidized good and $p_o = 2$ for the other good. Payoff functions and prices were the same in both stages. Unspent budget could neither be saved nor did it yield any payoff. There was no time limit for decisions. For these parameters, the consumption bundles (s, o) displayed in Figure 2.1 are as follows: the optimal consumption bundle in the reference stage is $A = (12, 7)$; the optimal bundle in the subsidy stage is $B = (13, 20)$; the bundle C is $(22, 7)$.

In order to make the difference between the initial endowment and the subsidy payment more salient, subjects had to earn their endowment in a real-effort task. Before consumption decisions were taken, subjects had to count the number of zeros in large spreadsheets that consisted of zeros and ones. When they managed to count

Consumed units	0	1	2	3	4	5	6	7	8	9	10	11	12
Payoff													
Subsidized good	0	36	70	102	132	160	186	210	232	252	270	286	299
Other good	0	30	57	81	102	120	135	147	157	166	175	184	192

Consumed units	13	14	15	16	17	18	19	20	21	22	23	24	25
Payoff													
Subsidized good	310	316	322	328	333	338	343	347	351	355	358	361	364
Other good	200	208	216	223	230	237	244	251	256	261	266	271	276

Table 2.1: *Payoff functions in the laboratory experiment. "Subsidized good" denotes the good that is subsidized in the second stage of the Lab Label treatment. Payoff points were converted into real money after the experiment.*

the correct number of zeros in a given amount of time they earned 100 money units that were later split in half for the two consumption decisions.[7] One subject failed to complete the task on time and is henceforth excluded from the analysis. We chose this rather boring activity to minimize the intrinsic motivation subjects could have for the task and thus to strengthen their perception of really having earned the money (cf. Cherry et al. 2002).

Subjects were students from the University of Bonn studying various majors except Economics. Treatments were assigned randomly and no subject participated in more than one treatment. At the beginning of the experiment, instructions were read aloud and subjects had to answer a number of control questions to ensure that they understood the task.[8] Detailed instructions for the two stages were given later on the computer screen. This allowed us to have subjects of both treatments in the same session and thus to align the delivery of the two treatments as much as possible. At the end of the experiment, subjects answered a questionnaire. The experiment was computerized using the software z-Tree (Fischbacher 2007). 92 sub-

[7]The precise rules were as follows: subjects got 8 large tables with 300 entries each. To complete the task, they had to count the correct number of zeros on four sheets within 15 minutes. An answer was also counted as correct if the number reported differed only by 1 from the true number. If subjects did not complete the task, they got an endowment of 10 money units only.

[8]See Appendix C.4 for an English translation of the instructions.

jects participated in the experiment, one of whom failed to complete the real-effort task. This leaves us with 45 independent observations in the Lab Cash treatment and 46 observations in the Lab Label treatment. Payoff points (cf. Table 2.1) were transformed into real money at a rate of 100 points = 1 euro. In addition to their earnings from the consumption decisions, subjects received a show-up fee of 2.50 euro. On average, subjects earned 12.20 euro (~14.80 USD at the time of the experiment). Sessions lasted between 60 and 70 minutes.

Results of the laboratory experiment are presented in Section 2.4.

2.3 Results of the Field Experiment

In this section, we report results of the field experiment. First we show that consumption of the subsidized good (beverages) is higher in the Field Label treatment than in the Field Cash treatment. Guests do not drink more in terms of volume but consume more expensive beverages. Total consumption, including meals, is not affected by the type of subsidy but spending is higher compared to when no voucher is given.

> **Result 1:** *Spending on the subsidized good (beverages) is significantly higher in the Field Label treatment than in the Field Cash treatment.*

Participants in the Cash treatment who receive a bill voucher spend on average 15.04 euro per capita on beverages (alcoholic and non-alcoholic). Participants in the Label treatment spend 18.94 euro on beverages. The treatment difference of 3.90 euro is very high considering the value of the subsidy (8 euro); the marginal effect is 25%. When we regress per-capita beverage consumption on a treatment dummy, the p-value is 0.052 (column 1 of Table 2.2). In a regression with controls for the share of women, share of persons below wine-drinking age, day of the week, and outside temperature, the treatment coefficient rises to 4.57 euro and the p-value goes down to 0.040 (column 2). The treatment difference is also significant in a non-parametric Mann-Whitney U-test ($p = 0.085$).[9]

[9]All tests reported in this chapter are two sided.

Unfortunately, we don't have information on one important predictor of consumption: total wealth (in terms of Figure 2.1, we cannot observe R). We pursue two alternative approaches to address this problem. First, we take *meal consumption* as proxy for total wealth: if richer people consume more beverages they should also consume more meals. We regress absolute beverage consumption per capita on a treatment dummy and control for meal consumption per capita. The coefficient of the treatment dummy rises to 4.90 euro and has now a p-value of 0.006 (column 3 of Table 2.2). This result holds when we add the controls mentioned above to the regression (column 4). A second possibility is to regress the *share* of beverage consumption on a treatment dummy and controls (columns 5 and 6 of Table 2.2). These regressions also show a highly significant impact of the treatment: giving a beverage voucher instead of a bill voucher increases the share of beverage consumption by 7.9 percentage points (from 33.7 percent; $p = 0.004$). In a regression with the controls described above (excluding meal consumption), the treatment effect goes up to 8.7 percentage points ($p = 0.004$). The difference in beverage shares is also significant in a U-test ($p = 0.013$). All these results and the results in the following regressions continue to hold if we drop the assumption that all observations are independent and instead only assume independence across days (by calculating robust standard errors clustered on day).[10]

Since the restaurant first distributed all beverage vouchers and then all bill vouchers, it might be that the treatment difference in beverage consumption is driven by an overall (falling) time trend. We can test this potential confound with the data we collected before and after the two main treatments; here, guests did not receive either voucher. If a time trend existed, it should also show up in this data. This is, however, not the case. Participants after the two main treatments spend even a little bit more on beverages (16.27 vs. 15.38 euro) but this difference is not significant, neither in a U-test ($p = 0.521$), nor in regressions with or without the controls described above[11] ($p = 0.863$ and $p = 0.562$). There is also no significant time trend

[10]The results also hold if we exclude tables that pay separately (only very few do so). The number of persons at a table also plays no role.

[11]Excluding share of women and share of kids, as we have this data only for observations *after* the two main treatments.

in total consumption or in the share of beverage consumption. [12]

Taken together, participants in our field experiment do not treat vouchers and wealth as completely fungible. Next, we analyze how the additional spending on the subsidized good in the Label treatment is used. Do participants consume a larger volume of beverages or do they consume more expensive beverages? Our data support the latter hypothesis.

> **Result 2:** *The volume of the consumed beverages does not depend on the type of subsidy given. By contrast, price per volume of the subsidized good is higher in the Field Label treatment.*

Participants in the Cash treatment drink on average 0.807 ltr of beverages (alcoholic and non-alcoholic) while participants in the Label treatment drink 0.857 ltr. The difference of 0.050 ltr (about 2 fl oz) is neither significant in a non-parametric U-test ($p = 0.420$) nor in OLS regressions with or without controls (see columns 1 and 2 of Table 2.3). In contrast, participants in the Label treatment consume more expensive beverages. The average price per liter is 21.91 euro/ltr in the Field Label treatment, while it is 18.39 euro/ltr in the Field Cash treatment. This difference of 3.52 euro is statistically significant (U-test, $p = 0.045$). In an OLS regression without controls, the p-value is 0.026; in a regression with the controls described above, the coefficient rises to 3.89 euro and the p-value is 0.022 (see columns 3 and 4 of Table 2.3).[13] We find the same pattern if we analyze the most expensive beverage item ordered. This item cost on average 12.20 euro in the Cash treatment and 19.93 euro in the Label treatment (U-test, $p = 0.003$).

So far we have argued that receiving a beverage voucher compared to a bill

[12] This result also speaks against a potential self-selection into the experiment because of word-of-mouth. It could be that early participants tell their neighbors about the vouchers and that these neighbors come to the restaurant to benefit from the vouchers. If this were true and if the neighbors had different consumption patterns, this could influence our treatment comparison. As restaurant guests did not know when the experiment ended, this would imply that we should also find an effect in the comparison of consumption behavior before and after the experiment. As shown above, there is no such difference.

[13] One table in the FCT did not consume any beverages and is excluded from this analysis. All other results stay the same if we exclude this observation also for those analyses.

Table 2.2: Treatment Effect of Receiving a Beverage vs. a Bill Voucher

	Absolute beverage consumption				Share of bev. consumption		Total consumption	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 if Label treatment	3.903*	4.573**	4.897***	5.109***	0.079***	0.087***	2.089	3.579
	(1.976)	(2.182)	(1.710)	(1.889)	(0.027)	(0.029)	(3.363)	(3.711)
Share of women		-1.617		0.333		-0.018		-5.233
		(4.541)		(3.946)		(0.061)		(7.724)
Share of kids		-10.764		-0.540		0.105		-29.723
		(14.636)		(12.829)		(0.197)		(24.893)
Outside temperature (in °C)		-0.391		-0.391		-0.005		-0.391
		(0.517)		(0.447)		(0.007)		(0.880)
Meal consumption			0.548***	0.539***				
			(0.109)	(0.113)				
Controls for day of the week	No	Yes	No	Yes	No	Yes	No	Yes
Constant	15.035***	21.116*	-0.122	5.988	0.337***	0.413***	42.712***	49.163**
	(1.426)	(10.903)	(3.250)	(9.943)	(0.019)	(0.147)	(2.428)	(18.544)
N.Obs.	71	71	71	71	71	71	71	71
Prob > F	0.052	0.346	0.000	0.000	0.004	0.083	0.537	0.731
Adjusted R^2	0.04	0.01	0.29	0.26	0.10	0.08	-0.01	-0.04

Table 2.2: *OLS estimates of the treatment effect on absolute and relative spending for the subsidized good (beverages) and on total spending. Notes: Standard errors are in parentheses. The sample includes observations from Field Cash and Field Label treatment. Absolute consumption is measured in euro. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.*

Table 2.3: Use of Increased Spending

	Consumed beverage volume		Price per liter	
	(1)	(2)	(3)	(4)
1 if Label treatment	0.050	0.086	3.520**	3.886**
	(0.072)	(0.081)	(1.541)	(1.653)
Share of women		-0.139		-0.167
		(0.169)		(3.424)
Share of kids		-0.025		-10.634
		(0.544)		(11.035)
Outside temperature (in °C)		0.005		-0.587
		(0.019)		(0.390)
Controls for day of the week	No	Yes	No	Yes
Constant	0.807***	0.710*	18.390***	27.903***
	(0.052)	(0.405)	(1.120)	(8.224)
N. Obs.	71	71	70	70
Prob > F	0.494	0.931	0.026	0.078
Adjusted R^2	-0.01	-0.06	0.06	0.08

Table 2.3: *OLS estimates of the treatment effect on volume (in liter) and price per volume (in euro/liter) of the subsidized good. Notes: Standard errors are in parentheses. The sample includes observations from Field Cash treatment and Field Label treatment. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.*

voucher should not alter consumption behavior if guests treated different income sources as fungible. The same argument can also be applied to the comparison of guests who receive either voucher and guests who do not receive such a subsidy. The 8 euro increase in lifetime income (by receiving the voucher) can surely be neglected. According to standard theory, we should not expect consumption to be influenced by receiving a voucher. Our data shows, however, that also in this comparison guests do not treat income sources as completely fungible.[14]

> **Result 3:** *Overall spending is higher in both subsidy treatments (FCT and FLT) compared to when no subsidy is given.*

Directly before and after the two main treatments we collected data of guests who did not receive a voucher. These guests spend on average 39.87 euro per capita. Guests in our two main treatments spend on average 3.93 euro more.[15] This difference is not significant in a U-test ($p = 0.157$) but in an OLS regression ($p = 0.059$, see Table 2.4, column 1). If we control for day of the week and outside temperature, the treatment effect rises to 4.92 euro and remains significant ($p = 0.080$, column 2).[16] Spending on beverages is also influenced by receiving a voucher. Since we expect a different impact on spending behavior according to the type of voucher, we regress absolute beverage consumption on two treatment dummies, one for the FLT and one for the FCT (Table 2.4, columns 3–6). Without controls, receiving a beverage voucher increases beverage consumption by 3.25 euro. This difference is significant

[14] Keep in mind that this effect could be influenced by a change in mood: when you get a gift you might be inclined to spend more, just because you are happy. The results so far and the results of the laboratory experiment cannot be influenced by mood.

[15] Average total consumption is not different between the two main treatments. In the Field Cash treatment, participants spend on average 42.71 euro per capita while participants in the Field Label treatment spend 44.80 euro. This difference is neither significant in a U-test ($p = 0.645$), nor in regressions with our without controls (see Table 2.2, columns 7 and 8). The amount of tip is also not different between the two main treatments and also not different compared to when no voucher is given (U-tests, FLT/FCT: $p = 0.416$, FLT/no voucher: $p = 0.327$, FCT/no voucher: $p = 0.777$). In Germany, there is no strong norm about how much tip to give; tipping in our data ranges from 0% to 12.5% of the amount of the bill.

[16] We do not have information about the share of women and the share of persons below wine-drinking age for observations before the experiment, thus we cannot control for these variables like we did in Table 2.2.

($p = 0.032$) and remains significant if we add the controls described above and/or if we control for meal consumption. The difference is also significant in a U-test ($p = 0.041$). In contrast, beverage consumption is not significantly higher when guests receive a bill voucher (see Table 2.4, columns 3–6; U-test, $p = 0.900$). The point estimate is even negative in some specifications. Apparently, guests who receive a bill voucher focus almost all additional spending on meals. This is also the reason why the share of beverage consumption is lower in the Cash treatment compared to guests without voucher (see Table 2.4, columns 7 and 8). In the Label treatment, the share of beverage consumption is higher compared to guests without voucher but not significantly. Our results on the comparison between guests with and without voucher are in line with findings of Milkman et al. (2007). They analyze data from an online grocer when customers redeem a $10-off coupon compared to when they don't. Controlling for customer fixed effects, they find that customers spend on average $1.87 more when they redeem such a subsidy.

Because of the high variance in total per-capita consumption, we could not exclude the possibility that some guests would like to consume less than the amount of the voucher, i.e., that their decisions are distorted. Indeed, in 16% of observations in the Field Cash treatment and of observations without voucher, absolute beverage consumption is lower than the value of the voucher. To avoid these distortions, we sort all observations by per-capita beverage consumption and exclude the lowest 16% in each treatment. In the resulting sample, the subsidy is inframarginal and non-distortionary for all participants. As shown in Tables 2.5 and 2.6, the distortions cause only a small part of the treatment differences and overall results remain very similar. The treatment difference between Label and Cash treatment in absolute beverage consumption is now between 3.40 and 4.53 euro depending on the exact specification (Table 2.5, columns 1–4). In the full sample, this effect was between 3.90 and 5.11 euro (see Table 2.2). The treatment effect remains significant (except for column 2). The share of beverage consumption increases by about 6.5 percentage points from Cash to Label treatment (full sample: 7.9 to 8.7 percentage points). We also obtain very similar results if we repeat the comparisons between guests who did or did not receive a voucher (see Tables 2.4 and 2.6): receiving a subsidy now

Table 2.4: Treatment Effect of Receiving Either Voucher

	Total consumption		Absolute beverage consumption				Share of bev. consumption	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 if Label or Cash treatment	3.930*	4.918*						
	(2.071)	(2.792)						
1 if Label treatment			3.251**	3.890**	2.507*	2.991*	0.025	0.033
			(1.509)	(1.842)	(1.353)	(1.650)	(0.023)	(0.027)
1 if Cash treatment			-0.652	0.092	-2.200	-1.521	-0.054**	-0.045
			(1.558)	(1.954)	(1.411)	(1.761)	(0.023)	(0.029)
Outside temperature (in °C)		-0.348		-0.239		-0.197		-0.002
		(0.358)		(0.209)		(0.187)		(0.003)
Meal consumption					0.443***	0.448***		
					(0.065)	(0.066)		
Controls for day of the week	No	Yes	No	Yes	No	Yes	No	Yes
Constant	39.871***	44.883***	15.687***	19.551***	4.971***	8.387**	0.390***	0.426***
	(1.276)	(5.772)	(0.742)	(3.380)	(1.695)	(3.434)	(0.011)	(0.050)
N. Obs.	187	187	187	187	187	187	187	187
Prob > F	0.059	0.417	0.065	0.403	0.000	0.000	0.018	0.095
Adjusted R^2	0.01	0.00	0.02	0.00	0.22	0.20	0.03	0.03

Table 2.4: *OLS estimates of the treatment effect compared to participants who do not receive a voucher. Notes: Standard errors are in parentheses. The sample includes observations from all three field treatments. Absolute consumption is measured in euro. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.*

increases consumption by 3.90 to 3.79 euro (full sample: 3.93 to 4.92 euro). The impact of a beverage voucher on beverage consumption compared to guest without voucher is now between 3.16 and 3.44 euro (full sample: 2.51 to 3.89 euro). Significance levels stay the same in most specifications.

Although it is reassuring to see that distorted participants do not drive the treatment difference in our field experiment, we would like to confirm the results in a setting where we can guarantee that the subsidy is inframarginal for all subjects. In a laboratory experiment, this is possible: we know budget and optimal decision and both are the same for all subjects.

2.4 Results of the Laboratory Experiment

In this section we report results from the laboratory experiment. First, we show that giving a labeled subsidy instead of a cash grant increases consumption of the subsidized good. Then we present evidence that this effect is stronger for subjects with lower mathematical abilities. Finally, we demonstrate that subjects' moral concerns cannot explain the treatment effect.

2.4.1 Consumption in the Experiment

The lab experiment consisted of two stages: in the *reference stage*, subjects allocated only a cash budget; in the subsequent *subsidy stage*, subjects had again the cash budget and an additional subsidy at their disposal. The total budget was identical for every subject. Before we turn to the subsidy stage, we analyze consumption decisions in the reference stage. The design of the reference stage was the same in both treatments. In particular, subjects were not aware of the fact that there would be two different treatments in the subsidy stage. Figure 2.2 shows a histogram of consumption choices for the (later to be) subsidized good. Choices in the Lab Cash treatment are represented by grey bars; choices in the Lab Label treatment are represented by black bars. We find that decisions are very similar: the modal choice in both treatments is the optimum of 12 units. On average, subjects in the Lab Label

Table 2.5: Treatment Effect of Receiving a Beverage vs. a Bill Voucher for Non-distorted Participants

	Absolute beverage consumption				Share of bev. consumption		Total consumption	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 if Label treatment	3.398*	3.463	4.528**	4.216**	0.066***	0.064**	0.854	1.735
	(1.936)	(2.224)	(1.746)	(2.007)	(0.024)	(0.028)	(3.420)	(3.896)
Share of women		-2.357		-0.247		-0.009		-7.195
		(5.156)		(4.662)		(0.064)		(9.031)
Share of kids		-12.950		-3.558		0.089		-34.483
		(13.249)		(12.160)		(0.165)		(23.206)
Outside temperature (in °C)		-0.544		-0.513		-0.008		-0.613
		(0.537)		(0.482)		(0.007)		(0.940)
Meal consumption			0.444***	0.436***				
			(0.111)	(0.118)				
Controls for day of the week	No	Yes	No	Yes	No	Yes	No	Yes
Constant	17.421***	28.890**	4.481	14.963	0.372***	0.536***	46.552***	60.820***
	(1.403)	(11.577)	(3.473)	(11.056)	(0.018)	(0.144)	(2.479)	(20.279)
N.Obs.	59	59	59	59	59	59	59	59
Prob > F	0.085	0.528	0.000	0.014	0.008	0.144	0.804	0.775
Adjusted R^2	0.03	-0.01	0.24	0.18	0.10	0.07	-0.02	-0.05

Table 2.5: *OLS estimates of the treatment effect on absolute and relative spending for the subsidized good. The sample is restricted to observations of non-distorted participants: we sort all observations by per-capita beverage consumption and exclude the lowest 15% in each treatment (see text). Notes: Standard errors are in parentheses. The sample includes observations from Field Cash and Field Label treatment. Absolute consumption is measured in euro. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.*

Table 2.6: Treatment Effect of Receiving Either Voucher for Non-distorted Participants

	Total consumption		Absolute beverage consumption				Share of bev. consumption	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 if Label or Cash treatment	3.900*	3.792						
	(2.029)	(2.783)						
1 if Label treatment			3.434**	3.436*	3.156**	3.253*	0.037*	0.038
			(1.511)	(1.838)	(1.428)	(1.730)	(0.021)	(0.025)
1 if Cash treatment			0.036	0.490	-1.054	-0.495	-0.029	-0.023
			(1.571)	(1.994)	(1.503)	(1.888)	(0.022)	(0.028)
Outside temperature (in °C)		-0.117		-0.146		-0.157		-0.003
		(0.354)		(0.210)		(0.198)		(0.003)
Meal consumption					0.319***	0.326***		
					(0.072)	(0.072)		
Controls for day of the week	No	Yes	No	Yes	No	Yes	No	Yes
Constant	43.100***	46.489***	17.384***	21.441***	9.176***	13.308***	0.401***	0.468***
	(1.244)	(5.722)	(0.741)	(3.409)	(1.975)	(3.677)	(0.010)	(0.047)
N. Obs.	157	157	157	157	157	157	157	157
Prob > F	0.056	0.558	0.069	0.260	0.000	0.000	0.046	0.086
Adjusted R^2	0.02	-0.01	0.02	0.01	0.13	0.13	0.03	0.04

Table 2.6: *OLS estimates of the treatment effect compared to participants who do not receive a voucher. The sample is restricted to observations of non-distorted participants: we sort all observations by per-capita beverage consumption and exclude the lowest 15% in each treatment (see text). Notes: Standard errors are in parentheses. The sample includes observations from all three field treatments. Absolute consumption is measured in euro. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.*

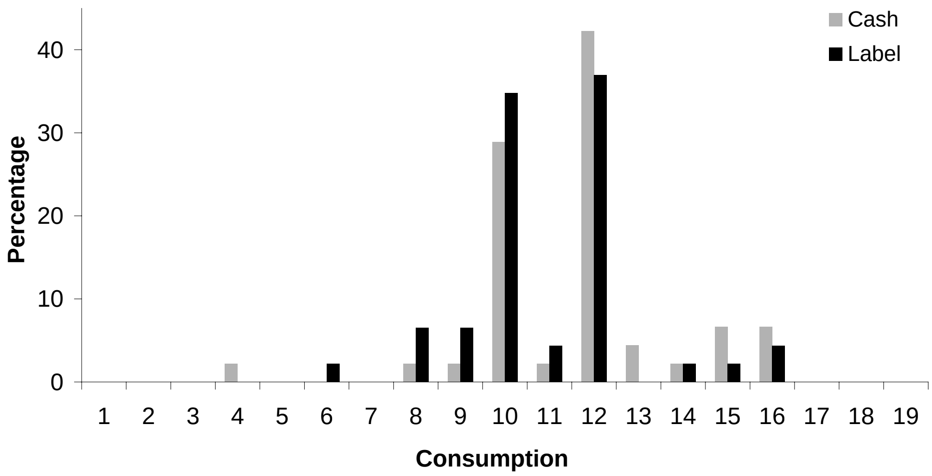


Figure 2.2: *Consumption of the subsidized good in the reference stage. A consumption of 12 units maximizes payoff.*

treatment consume 11.0 units of the subsidized good, while subjects in the Lab Cash treatment consume 11.6 units. Subjects apparently have no problem understanding the decision and take the decision seriously. We are therefore confident that the experimental situation allows for meaningful interpretation and that the experimental incentives work. OLS regressions show that consumption is not different between treatments (see Table 2.7: column 1 without controls, column 2 with controls for age and gender).[17]

Next, we focus on outcomes in the subsidy stage. Our first result concerns the impact of the labeled subsidy on consumption choice and confirms the main finding of the field experiment.

[17] A U-test between treatments is, however, significant ($p = 0.066$). Still, this does not compromise our results for several reasons: first, in terms of profits, there is virtually no treatment difference (U-test, $p = 0.983$). Second, the absolute distance to the optimum is not different between treatments (U-test, $p = 0.577$). Finally, consumption is slightly *lower* in the Lab Label treatment. Thus, if there is any inertia in subjects' decisions, results from the reference stage work against a potential treatment effect in the subsidy stage, making our results even stronger.

Result 4: *Consumption of the subsidized good is significantly higher in the Lab Label treatment. The marginal propensity to consume out of the subsidy is twice as large in the Lab Label treatment as in the Lab Cash treatment.*

A histogram of consumption choices in the subsidy stage is shown in Figure 2.3. Recall that the experiment is designed such that the same optimal consumption bundle can be reached in both treatments. If all subjects acted in line with fungibility, there should be no treatment difference. In the Cash treatment, we find that the modal choice is to consume the optimal amount of the subsidized good (13 units), and average consumption is 14.4 units.[18] In the Label treatment, by contrast, a consumption of an amount of 20 units is the most frequent choice and only a relatively small share of subjects choose the payoff-maximizing amount of 13 units. Overall, subjects in the Label treatment buy too much of the subsidized good, consuming 16.7 units on average; the marginal effect is 15%. The treatment difference is highly significant (U-test, $p = 0.006$). Moreover, subjects in the Label treatment leave money on the table, as their choices translate into significantly lower payoffs (U-test, $p = 0.014$). In Table 2.7 we provide OLS estimates of the treatment effect. The estimates in columns 3 and 4 show that the treatment dummy is positive and highly significant. The coefficient has a value of about 2.3 consumption units ($p = 0.006$) and remains unaffected when we control for age and gender.[19] To check the robustness of this finding we analyze an additional measure of performance: the absolute distance from the optimal consumption level. This measure also treats too low consumption as error. On average, subjects in the Cash treatment choose a consumption level that is 2.44 units away from the optimum; subjects in the Label treatment are on average 4.35 units away from the same optimum. The difference is significant (U-test, $p = 0.004$).

[18] Very few subjects in the subsidy stage choose a consumption bundle that is not on the Pareto frontier (one in the LLT and two in the LCT). For ease of exposition, we report only the consumption of the subsidized good. Consumption of the other good can then be readily calculated. Our results do not change if we confine the analysis to the Pareto optimal choices.

[19] All regression results in Table 2.7 remain virtually unchanged if we perform tobit regressions instead of OLS, controlling for the fact that subjects could not consume more than 25 units of each good.

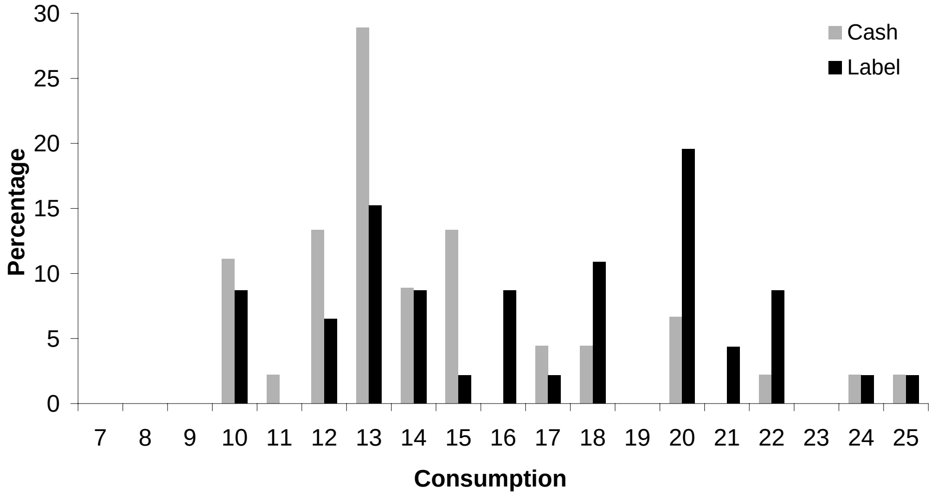


Figure 2.3: *Consumption of the subsidized good in the subsidy stage. A consumption of 13 units maximizes payoff.*

The two-stage design of our experiment enables us to compute an intra-person measure of fungibility by comparing decisions in the subsidy stage to decisions in the reference stage. A histogram of the intra-person change in consumption is shown in Figure 2.4. On average, the consumption increase in the Cash treatment is 2.8 units. In contrast, the average consumption increase in the Label treatment is 5.7 units. This difference-in-difference is highly significant (U-test, $p = 0.001$).[20] We can also express the consumption increase in terms of marginal propensity to consume the subsidized good (MPC). As the subsidy payment has a value of 10 units of the subsidized good and as the additional subsidy is the only budget change between the reference and the subsidy stage, the resulting MPC out of the subsidy is 0.574 in the Label treatment and 0.280 in the Cash treatment. Attaching a label to the subsidy therefore doubles the MPC out of the subsidy.

Our next result documents the considerable heterogeneity we observe in behavior across individuals.

[20] The increase from reference to subsidy stage of absolute distance to the optimal consumption level is also higher in the Label treatment (U-test, $p = 0.034$).

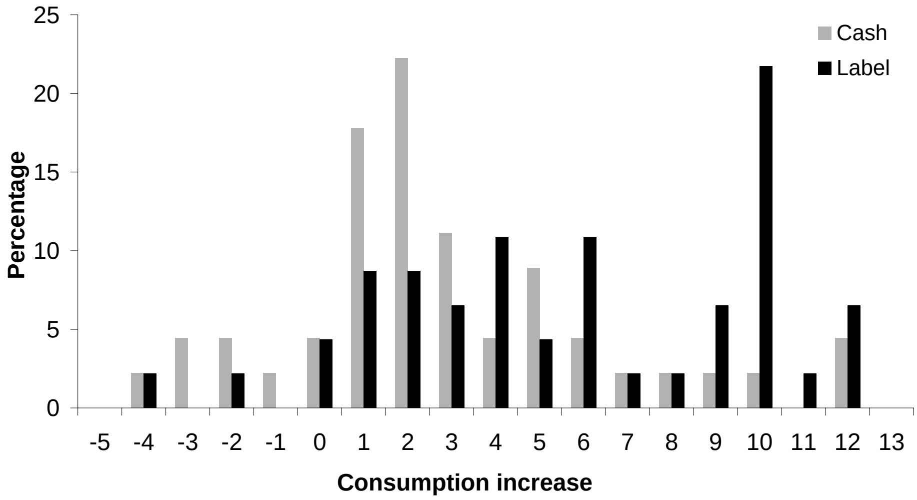


Figure 2.4: *Consumption increase of the subsidized good from reference stage to subsidy stage. The subsidy is worth 10 units of the subsidized good.*

> **Result 5:** *The treatment difference is mainly driven by subjects who increase their consumption by the full amount of the subsidy.*

Figure 2.4 shows that the most frequent choice in the Cash treatment is a consumption increase by either 1 or 2 units. In contrast, the modal choice in the Label treatment is a consumption increase by 10 units, i.e., these subjects spend the entire subsidy on the subsidized good, on top of the consumption from the reference stage. Subjects who treat income sources as completely non-fungible will do exactly this (cf. bundle C in Figure 2.1). In the Label treatment, 22% of subjects spend the whole subsidy on the subsidized good, while this is true for only 2% of subjects in the Cash treatment. These subjects drive most of the treatment effect, but not all of it. If we exclude these subjects from the analysis, the treatment difference in absolute consumption remains, although it is considerably smaller (1.4 units, previously 2.3 units; U-test, $p = 0.089$). The same is true for the treatment difference in consumption change (1.9 units, previously 2.9 units; U-test, $p = 0.020$).

Interestingly, subjects who spend the entire subsidy on the subsidized good decide much faster than the remaining subjects. They need on average 107 sec for their decision, whereas the other subjects need 234 sec, more than twice as long

(U-test, $p = 0.001$). This difference suggests that spending the subsidy fully on the subsidized good is the result of a simple decision heuristic (like mental accounting) rather than extensive deliberations. As a consequence of their consumption decision, subjects who spend the entire subsidy on the subsidized good earn less than all other subjects (U-test: $p = 0.001$) and also less than the other subjects in the Label treatment ($p = 0.003$).

2.4.2 Impact of Mathematical Abilities

A consumer who does not treat different income components as fungible reduces the complexity of the consumption decision. In our setup, ignoring fungibility divides the rather complex two-good decision into two simple one-dimensional problems. Subjects with lower mathematical skills will have a larger gain from reducing the complexity of the decision. We therefore expect these subjects to violate fungibility more often and, as a consequence, to be more influenced by the treatment manipulation. We mentioned in Section 3.1 that a consumer who brackets his decisions narrowly, i.e., who does not decide globally, will violate fungibility. Read et al. (1999*b*) conjecture that narrow bracketing is negatively correlated with cognitive and mathematical abilities.[21] Our next result supports their intuition. To analyze the interplay of cognitive abilities and the treatment effect, we use subjects' math grade in their final high school exam as a proxy for their cognitive and mathematical abilities.

> **Result 6:** *The treatment difference in consumption is larger for subjects with lower mathematical abilities.*

The grades were elicited in the post-experimental questionnaire. We split the sample at the median according to math grade, leading to a "High-Math" group (n= 50) and a "Low-Math" group (n=41). Since mathematical abilities might not only influence the decision in the Label treatment but also in the Cash treatment, we compare

[21] Thaler (1985) argues that mental accounting, a concept similar to narrow bracketing, serves as a heuristic to overcome problems of limited self-control. In our experiment, limited self-control plays no role. Subjects with lower cognitive abilities should be less likely to realize this fact and should thus be more likely to use this heuristic.

Table 2.7: Consumption in the Laboratory Experiment

	Consumption of the subsidized good							
	Reference stage		Subsidy stage					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 if Label Treatment	-0.666	-0.652	2.273***	2.270***	0.911	0.869	2.091*	2.088*
	(0.438)	(0.439)	(0.808)	(0.818)	(1.081)	(1.103)	(1.069)	(1.081)
1 if Low-Math Group					-0.796	-0.861		
					(1.160)	(1.191)		
Label * Low-Math					2.883*	2.960*		
					(1.615)	(1.653)		
1 if High-MO Group							-0.970	-0.989
							(1.157)	(1.186)
Label * High-MO							0.240	0.254
							(1.655)	(1.676)
Age		-0.077		0.021		0.047		-0.010
		(0.075)		(0.140)		(0.141)		(0.144)
1 if Female		-0.402		0.132		0.068		0.115
		(0.458)		(0.852)		(0.843)		(0.858)
Constant	11.622***	13.605***	14.422***	13.870 ***	14.741***	13.663***	14.875***	15.045***
	(0.311)	(1.769)	(0.574)	(3.293)	(0.733)	(3.259)	(0.790)	(3.511)
N. Obs.	91	91	91	91	91	91	91	91
Prob > F	0.132	0.270	0.006	0.057	0.010	0.046	0.036	0.132
Adjusted R^2	0.01	0.01	0.07	0.05	0.09	0.07	0.06	0.04

Table 2.7: *OLS estimates of the treatment effect on consumption of the subsidized good in the laboratory experiment. Notes: Standard errors are in parentheses. The dependent variable is consumption of the subsidized good in the reference stage (columns 1 and 2) or in the subsidy stage (columns 3 to 8) measured in units. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.*

the treatment difference within the High-Math group to the difference within the Low-Math group.

Figure 2.5 presents cumulative percentages of consumption decisions in both treatments. In the High-Math group (left panel), cumulative distributions are very close to each other. In contrast, in the Low-Math group (right panel), there is a wide gap between the two distributions. The treatment effect in the High-Math group is 0.9 units (U-test, $p = 0.357$). The treatment effect in the Low-Math group of 3.8 units is considerably larger (U-test, $p = 0.004$). We test whether the difference between treatment effects is statistically significant in OLS regressions. Results are presented in Table 2.7, columns 5 and 6. The dependent variable is the consumption of the subsidized good. As explanatory variables we include a treatment dummy (=1 if Label treatment), a Math dummy (=1 if Low-Math) and the interaction of the two. Compared to regressions without controls for mathematical abilities (columns 3 and 4), the treatment dummy is much smaller and insignificant. This indicates that subjects in the High-Math group (the reference category) do not act differently across treatments. In contrast, the coefficient of the interaction term is large (2.9 units) and significant. Controlling for age and gender leaves the significance level unaffected, and the coefficient increases slightly. Thus, the treatment effect is significantly larger for subjects who belong to the Low-Math group. We obtain the same result if we ignore the Cash treatment and directly compare consumption of High-Math and Low-Math subjects in the Label treatment. Low-Math subjects consume on average 2.1 units more of the subsidized good than High-Math subjects (U-test, $p = 0.066$). All these results hold if we take absolute distance to the optimal consumption as dependent variable.[22]

We have shown before that subjects who spend the whole subsidy on the subsidized good account for a large part of the treatment effect. The relation we have just shown between mathematical abilities and consumption in the subsidy stage holds also for these subjects. 90% of subjects who spend the entire subsidy on the subsidized good belong to the Low-Math group. Among all other subjects, this

[22] In the German high school system, there are two types of math course: intensive and basic course (*Leistungskurs* and *Grundkurs*). If we control for subjects' course results are similar.

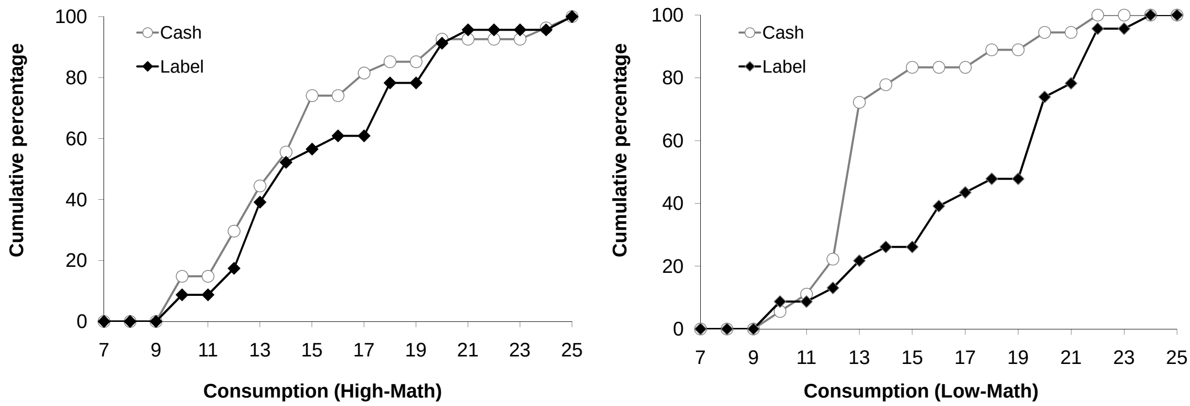


Figure 2.5: *Cumulative distribution of the consumption of the subsidized good for High-Math group (left panel) and Low-Math group (right panel). The grey line always depicts consumption in the Cash treatment, the black line consumption in the Label treatment. In order to form a High-Math group (n=50) and a Low-Math group (n=41) we elicited the math grades subjects obtained in their final high school exam and split the sample at the median.*

share is only 39.5%.

2.4.3 Moral Obligation as an Alternative Explanation?

So far we have attributed the treatment difference to cognitive limitations that prevent subjects from treating the two income components as fungible. However, one could also imagine that receiving a benefit payment causes a feeling of moral obligation to spend the money in accordance with the benefit giver's intention. In response to the intention that is conveyed by the label, recipients might increase their consumption of the subsidized good above the level they would have chosen if they had received the same amount as an unconditional cash payment. This reasoning cannot explain the behavioral patterns according to mathematical ability that we presented in the last section (unless one makes the rather strong assumption that mathematical abilities and moral attitudes are negatively correlated). But perhaps a feeling of moral obligation could be an additional factor influencing behavior. The next result shows, however, that behavior in the experiment is not related to subjects' attitudes towards moral obligation.

Result 7: *A feeling of moral obligation to comply with the label on the subsidy cannot explain the treatment difference in consumption.*

To measure subjects' attitude regarding moral obligation, we included a scenario in the post-experimental questionnaire in which subjects had to judge the behavior of a fictitious person. The scenario reads as follows (translated from German):

> Mr and Mrs Miller have two children (5 and 8 years old). They earn a total amount of 2000 euro per month, after taxes. Additionally, they receive 180 euro child benefit per child, i.e., a total of 360 euro per month. Usually, they spend about 300 euro per month for their children (child clothing, toys, etc.). They spend the rest of the child benefit on other things (e.g., their own hobbies).

Subjects had to indicate on a point scale from 1 to 6 how they judge the fictitious persons' behavior, 1 indicating "not appropriate at all" and 6 indicating "completely appropriate". Thus, a higher number indicates a *weaker* feeling of moral obligation. The decision situation described in the scenario above is very similar to the consumption decision in our experiment. In both situations, the intended use of the subsidy is obvious but the subsidy is not binding, i.e., a rational decision maker should not be influenced by the label attached to the subsidy.[23]

Analogous to our analysis of mathematical abilities, we split the sample at the median. This results in a group with strong moral obligation ("High-MO", n=38) and a group with weak moral obligation ("Low-MO", n=53). The cumulative percentages for both groups are shown in Figure 2.6. In both panels, the cumulative distributions are equally far from each other, indicating that the effect of the label does not depend on the moral concerns of the subjects. The treatment effect is essentially the same in both High-MO and Low-MO group (2.3 vs. 2.1 units); the significance levels of U-tests within each group are $p = 0.119$ and $p = 0.057$. This

[23]The questionnaire included two other scenarios concerning (i) a person claiming student support provided by the state in spite of not being entitled to it, and (ii) a person temporarily claiming unemployment benefits although having a new job already on the horizon. The full texts of these two scenarios are reported in Appendix B.3. The results for all three scenarios are very similar.

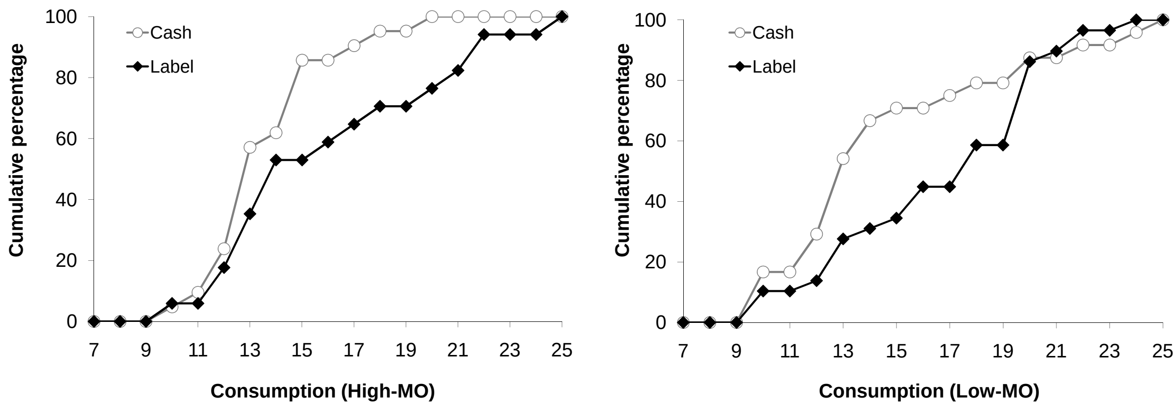


Figure 2.6: *Cumulative distribution of the consumption of the subsidized good for High-Moral-Obligation group (left panel) and Low-Moral-Obligation group (right panel). The grey line depicts consumption in the Cash treatment, the black line consumption in the Label treatment. In order to form a High-MO group (n=38) and a Low-MO group (n=53) we elicited moral attitudes via a vignette question concerning the use of child benefits and split the sample at the median.*

result is confirmed by a regression analysis. Columns 7 and 8 of Table 2.7 include a dummy equal to one if a subject is in the High-MO group and an interaction term between the treatment dummy and the High-MO dummy. We find that, in contrast to the math regressions in columns 5 and 6, the coefficient on the treatment dummy remains large (2.1 units) and significant. This indicates that also subjects with weak moral obligation are influenced by the label on the subsidy. Moreover, the interaction term is small (0.2 units) and far from being significant. Thus, the group of subjects with stronger moral obligation does not show a larger treatment effect.[24] This holds true if we focus on the subjects who spend the whole subsidy on the subsidized good in addition to their consumption in the reference stage. These subjects do not differ in their moral obligation from other subjects (U-test, $p = 0.298$). Taken together, these results indicate that moral obligation does not drive the treatment effect in the lab experiment.

[24]The answers to the two other vignettes are also not systematically related to the consumption of the subsidized good. In a similar regression as in column 7 of Table 2.7, the interaction terms are +0.91 units and -1.07 units for the student-support scenario and the unemployment-benefit scenario, respectively. Both are not significant. The same holds true if we take the average answer of each subject to the three vignettes as our proxy for moral obligation.

2.5 Discussion and Conclusion

In this chapter we pursued a dual research strategy by combining a natural field experiment and an incentivized laboratory experiment to test whether consumers treat different income sources as fungible. Both experiments yield the same results: even in a simple setup, many subjects do not act in line with fungibility. This effect is stronger for persons with lower mathematical skills. Differences in preferences, e.g, concerning the moral obligation to comply with the intention of the subsidy giver, do not drive our results.

As discussed in the introduction, our findings have implications for the study of stock market behavior and life-cycle savings. We furthermore suggest that fungibility plays an important role in a different setting where it has until now not been considered: the effect of in-kind benefits on consumption and market prices. Empirical studies for the U.S. and for France have shown that a rise in housing benefits for low-income tenants has lead to drastic rent increases (see, e.g., Susin 2002, Fack 2006).[25] For the U.S., where not all eligible households do receive the benefits, Susin (2002) finds that housing benefits even caused a *net loss* for low-income households. The standard explanation for this phenomenon is that the subsidy causes an increase in housing demand which is met by an inelastic supply. Our findings suggest that this is only part of the story. Taking our results at face value, tenants who receive housing benefits will have a higher willingness to pay for a given apartment, compared to tenants who receive the same amount as a cash grant. Landlords who anticipate this effect can increase the rent accordingly. Laferrère & Le Blanc (2004) present evidence from France that supports this view: controlling for apartment and neighborhood characteristics, landlords discriminate between non-assisted tenants and tenants who receive housing assistance, charging the latter group significantly higher rents. As a result, housing benefits do not necessarily make the recipients better off, but may constitute a transfer payment from taxpayers to landlords. In our view, this problem can be mitigated by linking housing benefits less saliently

[25]Similarly, Gibbons & Manning (2006) show for the U.K. that a reduction in housing benefits has lead to lower rents.

to rent payments. The periodicity of the benefit payments, for instance, could be chosen such that it differs from the periodicity of the rent payments. Moreover, one could design the benefit system such that the exact amount of the subsidy depends on variables which the landlord cannot observe.

There are, however, other benefit payments for which it is good news if recipients violate fungibility. If recipients of child benefits do not treat money as fungible, they will spend more of the subsidy on child-related goods, compared to a situation in which they receive a cash grant of the same amount.[26] This effect has been documented with data from the Dutch child benefits system: Kooreman (2000) finds that the marginal propensity to consume child clothing out of child benefits is higher than it is out of other income.[27] For a similar benefit system, Munro (2005) finds that the unconditional "winter fuel allowance" in the UK has a positive effect on heating expenditures. These results suggest that public policy can influence consumers in a simple way by explicitly stating the intended use of the subsidy (Thaler & Sunstein 2003).

Our results do not imply that everybody in every situation will violate fungibility. In our laboratory experiment, treating money as non-fungible is linked to mathematical abilities and not to preferences, suggesting that this behavior is a mistake. Once the rational solution becomes obvious to subjects, e.g., by learning or by explanation, they will probably regret their decision and choose the optimal solution. The field experiment shows, however, that also experienced participants can be influenced by a label on the subsidy. In addition, coming back to our previous example of housing benefits, most people make only few housing decisions in their life; here, the scope for non-rational behavior should be especially large.

[26]Recipients of child benefits are not restricted in their use of these funds and only the name of the benefit payment marks it as a separate income component. But the general argument of this chapter applies if we assume that also merely labeled payments are posted to the mental sub-budget corresponding to the label.

[27]Blow et al. (2004), however, analyze data from the U.K. and find such a labeling effect only for some of their specifications.

Chapter 3

Reference-Dependent Preferences and Labor Supply

3.1 Introduction

People often judge the outcome of a decision not only according to their intrinsic taste for the outcome but also in comparison to a reference point. This has been demonstrated in many different setups including investment and consumption decisions, labor supply, and self-reported happiness (e.g., Kahneman & Tversky 1979, Easterlin 1974, Camerer et al. 1997, Odean 1998, Genesove & Mayer 2001). A key open question, however, is what people take as the reference point for their decisions. Answering this question and developing an empirically validated theory of the reference point is of crucial importance to discipline predictions. Otherwise, if the reference point is assumed case-by-case, reference-dependent preferences (RDP) models might explain behavior not because of their structural assumptions but because of this additional degree of freedom.

The original formulation of RDP theory by Kahneman & Tversky (1979) is remarkably silent on what constitutes a reference point. Most empirical studies assume the status quo or lagged status quo as reference point, and have shown that behavior is consistent with such a reference point (e.g., Thaler & Johnson 1990, Kahneman et al. 1990). Other reference points, however, might be a better predictor of behavior. One promising candidate is an individual's rational expectation. In many situations, assuming expectations instead of status quo as reference point leads to intuitively more plausible predictions. Imagine a worker who expects a salary increase of 10 percent but only gets an increase of 5 percent. This will be felt as a loss by most people, although it is a gain compared to the status quo. Other examples where expectations seem to serve as a reference point include deciding whether to buy a good with an unexpectedly high price (Heidhues & Kőszegi forthcoming), not getting an item on eBay after temporarily having been the highest bidder, or being rejected after a "revise and resubmit" compared to being rejected directly.

The major problem for empirical tests of RDP models is that reference points cannot be observed directly. Testing the importance of a potential reference point thus ideally requires an exogenous manipulation of this reference point in a tightly controlled environment. In this chapter, we thus report evidence from a real-effort experiment in which we exogenously vary expectations; we show that behavior is systematically and causally affected by this exogenous variation.

The design of the experiment is guided by a class of RDP theories (Bell 1985, Loomes & Sugden 1986, Gul 1991, Kőszegi & Rabin 2006, Kőszegi & Rabin 2007) which assumes that expectations serve as reference point. This makes it possible to derive clear-cut theoretical predictions and to clearly distinguish between alternative hypotheses. The experimental design creates a situation in which expectation-based RDP models give sharply different predictions compared to status-quo models of RDP or a standard model of labor supply in the canonical expected utility framework. The experiment thus provides a test between important classes of models. A key advantage of the experiment is that rational expectations of participants are known to the researcher, something which is difficult to achieve in the field.

In the experiment, subjects work on a tedious, repetitive task. After each repetition, they can decide whether to continue or to stop working. They get a piece rate, but receive their accumulated piece rate earnings only with 50 percent probability, whereas with 50 percent probability they receive a fixed payment instead. Which payment subjects receive is determined only after they have made their choice about when to stop working. The only treatment manipulation is a variation of the amount of the fixed payment.

The standard model with separable utility over money and effort costs does not predict a treatment difference. Effort is determined by setting marginal cost equal to the marginal benefit defined by the piece rate, and the fixed payment is irrelevant for either marginal cost or marginal benefit of effort. This is true independent of the shape of utility over money, and the shape of the cost function. Models in which the status quo is the reference point also predict no treatment difference, because the status quo when entering the experiment is the same for both treatments.

A model with expectation-based reference points such as Kőszegi & Rabin (2007) predicts higher effort in the treatment with the higher fixed payment. This is because the treatment manipulation changes expected payoffs, which affects an individual's expectation about whether a given effort choice will feel like a loss or a gain once uncertainty has been resolved. In particular, as long as the accumulated piece rate is below the fixed payment, the individual risks getting the accumulated piece rate instead of the (higher) fixed payment, which would feel like a loss. This expectation induces the worker to exert more effort. Once accumulated earnings are higher than the fixed payment, however, getting the fixed payment feels like a loss, and working harder only magnifies the size of the loss. This expectation induces lower effort. The model thus predicts that the propensity to stop is higher when the accumulated piece rate equals the fixed payment, and that increasing the size of the fixed payment will tend to increase overall effort.

We find a strong treatment effect, such that individuals in the high fixed payment treatment work significantly more. The size of the increase in labor supply is large relative to the treatment manipulation: average earnings increase by about 2 euros in response to a 4 euro difference in the fixed payment amount across treatments. We also observe strong spikes in the distribution of effort choices, exactly at the low fixed payment amount in the low fixed payment treatment, and at the high fixed payment amount in the high fixed payment treatment. Moreover, there is no spike at the high fixed payment amount in the low fixed payment treatment, and vice versa. These findings are consistent with the main predictions of the RDP models with reference points in expectations. We also find support for a more subtle prediction of these models, which is that the treatment difference should be driven by individuals who stop in the interval between the low and high fixed payment amounts. Additionally, when asked in a questionnaire after the experiment why they stopped, almost all subjects stopping at the fixed payment gave reasons like that they wanted to avoid "uncertainty" or to prevent "losing their earnings" beyond the fixed payment. These answers suggest that loss aversion or disappointment aversion drive our results. Our findings do not conflict with previous evidence on the relevance of status quo reference points for decision making, given that in previous studies the status quo has typically coincided with expectations.

Related evidence comes from the literature on violations of expected utility theory in lottery choices, in which some findings are also supportive of a role for expectation-based reference points in decision making (see Bell (1985), and Loomes & Sugden (1986, 1987) for discussions). Different from this chapter, the evidence has mainly come from inconsistencies observed in choices involving relatively complex combinations of different financial lotteries. Recent evidence from Rabin & Weizsäcker (2007) raises the possibility that some violations of expected utility theory observed in complex lottery choices may go away, if cognitive costs are lowered. For example they find that working out the math for subjects eliminates the particular violation known as narrow bracketing violation. Our experiment is complementary in that it measures the impact of reference points as expectations in the domain of real effort choices, rather than lottery decisions. Moreover, it provides corroborating evidence on the importance of reference points as expectations, based on a very simple and transparent test, where subjects can act in accordance with expected utility theory simply by ignoring the fixed payment.

Besides a contribution to the fundamental question what constitutes a reference point, our results are also relevant for the lively debate regarding labor supply and transitory wage changes. It is debated whether the response of labor supply to changes in incentives is consistent with the standard intertemporal substitution of labor and leisure or rather with loss aversion around a daily reference income (e.g., Camerer et al. 1997, Farber 2004, Farber 2005, Fehr & Götte 2007). A key lacuna in this literature is that the reference point is unobserved. Our experiment makes the reference point known to the researcher. We discuss this issue in more detail in Section 3.5.

The chapter is organized as follows. Details of the experimental design are explained in the following section. Section 3.3 discusses behavioral predictions.

Results are presented in Section 3.4. Section 3.5 concludes.

3.2 Design

In the experiment, subjects work on a tedious task: they count the number of zeros in large tables that consist of zeros and ones. Subjects get a piece rate per correctly counted table; then a new table is generated.[1] We chose this task because no prior knowledge is required, performance is perfectly measurable, and there is little learning possibility; at the same time, the task is boring and pointless and we can thus be confident that the task entails a positive cost of effort. In the first of two stages, subjects have four minutes to count as many tables as possible. They get a piece rate of 10 cents per correct answer for sure. This part serves to familiarize subjects with the task; additionally, we will use performance in this stage as a productivity indicator. In the second and main stage of the experiment, subjects do the same task again—counting zeros—but they can now decide themselves how much and for how long they want to work. At most, they can work for 60 minutes on the task. When they want to stop, they can push a button on the screen and the experiment is over: subjects get paid immediately and can leave.

To exogenously vary expectations, subjects do not get their accumulated piece rates of the main stage for sure. Before they start counting in this stage, they have to choose one of two closed envelopes. They know that one of the envelopes contains a card "You get the amount you worked for." and that the other envelope contains a card "You get 3 euros." But they do not know which card is in which envelope. The envelopes remain with the subject while he is working and are only opened after the subject has stopped working. The subject's payment is then determined by the card in their chosen envelope. To keep economic incentives comparable between the two stages, the piece rate per correct answer is doubled to 20 cents in the main stage. The only difference between the two treatments is the amount of the fixed

[1]Each table contains 150 entries. If an answer is not correct, subjects have two more tries for the same table. To prevent guessing, the piece rate is *deducted* from their account if they fail all three tries. This happened only 26 times in the experiment (vs. more than 5000 correctly counted tables).

payment. In the LO-treatment, the fixed payment is 3 euros while it is 7 euros in the HI-treatment.[2]

Because the uncertainty about the payment is revealed only after the work is finished, we know the rational expectation of each subject when deciding whether to stop working: with 50 percent probability the subject gets the amount he worked for and with 50 percent he gets the fixed payment of 3 or 7 euros.

A potential confound could arise if subjects start working simultaneously and work in the same room, e.g., due to peer effects (Falk & Ichino 2006) or due to a desire for conformity (Bernheim 1994). We employed a particular procedure to prevent such effects: subjects came one by one, individual starting times were at least 20 minutes apart. Upon arrival, subjects were guided to one of three identical, neutral rooms.[3] They worked alone in their room with the door closed and never (with very few exceptions) saw another subject or the other two experimental rooms. Instructions and payments were also administered in their room. Because of this special procedure, subjects' stopping behavior cannot be influenced by other subjects' behavior in a systematic way.

Subjects were students from the University of Bonn studying various majors except Economics. We recruited subjects who had participated in no or only few previous experiments to avoid strong expectations about earnings from the experiment. Experiments were computerized using the software z-Tree (Fischbacher 2007). Treatments were assigned randomly and no subject participated in more than one treatment. We also randomized treatments over morning and afternoon time slots and over days of the week. 120 subjects participated in the experiment, 60 in each treatment. In addition to their earnings from the two stages of the experiment, subjects received a show-up fee of 5 euros. On average, subjects earned 13.70 euros (about 18.40 USD at the time of the experiment). The experiment took about one hour on average.

[2]For an English translation of the instructions, see Appendix C.4.

[3]Photos of the three rooms are shown in Appendix C.5.

3.3 Predictions

We examine in turn the predictions for our experimental setup of three theoretical models: no reference dependence, status-quo reference dependence, and expectation-based reference dependence.

First, consider a standard model of labor supply with a utility function separable in monetary payoff and cost of effort: $U(x, e) = u(x) - c(e)$. x is the monetary payoff and $c(e)$ is the cost of effort with $\partial c/\partial e > 0$. In our experimental setup, this utility function becomes $U(e, f, w) = \frac{1}{2}u(f) + \frac{1}{2}u(we) - c(e)$. With probability $\frac{1}{2}$ the subject gets the fixed payment f and with probability $\frac{1}{2}$ he gets the accumulated earnings we, with $w > 0$ the piece rate per table and e the number of correctly solved tables. In either case, he has to bear the cost of effort $c(e)$. The subject decides after each table whether to continue or to stop, yielding the following first-order condition:

$$\frac{\partial U}{\partial e} = \frac{w}{2}u'(we) - c'(e) \quad \Rightarrow \quad u'(we^*) = \frac{2}{w}c'(e^*)$$

The optimal effort level e^* is independent of the fixed payment f because in the state of the world when the subject gets the fixed payment f, he wants to stop right away, independent of the size of the fixed payment. And in the state of the world when he gets the accumulated earnings, the fixed payment does not matter either. This result holds regardless of whether the subject is risk-neutral, risk-averse, or risk-loving. The shape of the cost function has also no influence (as long as it increases).

We should also expect no treatment difference in labor supply if subjects have reference-dependent preferences with the reference point being the status quo. A crucial feature of reference dependent preferences is *loss aversion*: an outcome that is lower than the reference point is felt as a loss and weighs more heavily than an equal sized gain (Kahneman & Tversky 1979). Loss aversion with a status-quo reference point can not cause a treatment difference either because the status quo at the beginning of the experiment is the same in both treatments and independent of f.

In contrast to behavior predicted by the two previous models, subjects with expectation-based reference-dependent preferences (e.g., Bell 1985, Kőszegi & Rabin 2006) should work harder when the fixed payment f is higher. In these models, individuals compare their outcome to their expectation and feel a loss if the outcome falls short of their expectation. We calculate the predictions of the model of Kőszegi & Rabin (2006), but the models by Loomes & Sugden (1986), Gul (1991), and Bell (1985) make similar predictions.[4]

In the model of Kőszegi & Rabin (2006), an individual's utility depends on the K-dimensional consumption bundle c and on a reference bundle r. He has intrinsic consumption utility $m(c)$ but overall utility is given by $u(c|r) = m(c)+n(c|r)$, where $n(c|r)$ is "gain-loss utility". Both consumption utility and gain-loss utility are separable across dimensions, so that $m(c) = \sum_k m_k(c_k)$ and $n(c|r) = \sum_k n_k(c_k|r_k)$. The gain-loss utility is defined as $n_k(c_k|r_k) = \mu(m_k(c_k) - m_k(r_k))$, i.e., μ of the difference between the consumption utility of outcome and reference point, where $\mu(s)$ generally satisfies the properties of the value function of Kahneman & Tversky (1979). For small arguments s, Kőszegi & Rabin assume that $\mu(s)$ is piece-wise linear: $\mu(s) = \eta s$ for $s \geq 0$ and $\mu(s) = \eta\lambda s$ for $s < 0$ with $\eta \geq 0$ and $\lambda > 1$. The fact that λ is strictly greater than 1 captures loss aversion: losses loom larger than equal-sized gains. The model allows for both stochastic outcomes and stochastic reference points, and assumes that a stochastic outcome F is evaluated according to its expected utility, with the utility of each outcome being the average of how it feels relative to each possible realization of the reference point G: $U(F|G) = \int\int u(c|r)dG(r)dF(c)$. The reference point G is the probabilistic beliefs the individual held in the recent past about outcomes. The beliefs are generated by rational expectations.

In the setup we study in this chapter, the outcomes are not very large and we therefore assume that gain-loss utility is piece-wise linear as described above and that also consumption utility is linear: $m(c) = c$. Since we conduct a real effort experiment, it is natural to assume that subjects assess outcomes along $K = 2$ dimensions: money and effort/leisure. We calculate the choice-acclimated personal equilibrium (Kőszegi & Rabin 2007) since subjects had time to adjust their expec-

[4]The main difference between Kőszegi & Rabin (2006) and the other theories is how expectations are mapped into a reference point. Bell (1985), for example, assumes it to be the mean while Kőszegi & Rabin (2006) assume that each outcome is compared to the whole expectation distribution, but this distinction does not matter for our setup.

tations when they make the decision to stop or to continue.

This translates into the following utility functions: if current accumulated earnings are below the fixed payment ($we < f$) and the subject decides to stop, he gets

$$U = \frac{we + f}{2} - c(e) + \frac{1}{4}\eta(f - we) + \frac{1}{4}\eta\lambda(we - f)$$

The first two terms are the expected earnings of getting we or f and the cost of effort the subject has to bear in either case. The last two terms describe the gain-loss utility where outcomes are compared to rational expectations. Because of the experimental design, we know exactly the rational expectations of a subject when he decides whether to stop: with probability $\frac{1}{2}$ he expects to get the accumulated earnings and with probability $\frac{1}{2}$ he expects to get the fixed payment. For each belief, both outcomes have probability $\frac{1}{2}$, thus each belief-outcome combination occurs with probability $\frac{1}{4}$. Getting the fixed payment f when expecting the (lower) we feels like a gain (third term). As described above, the gain-loss utility is η times the distance between outcome and expectation. Getting we when expecting to get f feels like a loss. This fourth term is thus weighted with $\lambda > 1$. In two of the four belief-outcome combinations, the subject expects the outcome correctly and there is no gain-loss utility. The cost of effort never shows up in the gain-loss utility as the subject decides after each table whether to continue or to stop working. Thus the expectation about effort and cost of effort is always correct; there is only uncertainty about the monetary payoff.

If the current accumulated earnings are higher than the fixed payment ($we \geq f$), the gain-loss utility changes. Getting the accumulated earnings now feels like a gain compared to the lower fixed payment (third term), while getting the fixed payment now means a loss (fourth term):

$$U = \frac{we + f}{2} - c(e) + \frac{1}{4}\eta(we - f) + \frac{1}{4}\eta\lambda(f - we)$$

The first-order conditions are then:

$$we < f: \quad \frac{\partial U}{\partial e} = \frac{w}{2} - c'(e) + \frac{1}{4}\eta(\lambda - 1)w \quad \Rightarrow \quad c'(e^*) = \frac{w}{2} + \frac{w}{4}\eta(\lambda - 1)$$

$$we \geq f: \quad \frac{\partial U}{\partial e} = \frac{w}{2} - c'(e) - \frac{1}{4}\eta(\lambda - 1)w \quad \Rightarrow \quad c'(e^*) = \frac{w}{2} - \frac{w}{4}\eta(\lambda - 1)$$

If accumulated earnings are below f, the gain-loss utility makes effort *less* costly as stopping now entails a loss in case the subject gets we. If the accumulated earnings are above f, loss aversion makes effort *more* costly as effort beyond f can be lost in case he gets the fixed payment f.

How would increasing the fixed payment from f_{LO} to f_{HI} influence the labor supply decision with this utility function? It increases effort for all those individuals who for $f = f_{LO}$ would stop at accumulated earnings we between f_{LO} and f_{HI}. For these individuals, the gain-loss utility that before held back their effort, now (with $f = f_{HI}$) induces them to exert more effort. They can now work harder without risking to feel a loss when getting the fixed payment. If for $f = f_{LO}$ the individual would stop at an accumulated earnings level below f_{LO} or above f_{HI}, this change of f cannot have an effect. The gain-loss utility does not depend on the distance between we and f. All that matters is whether we is larger or smaller than f. Increasing f from f_{LO} to f_{HI} does not change this relation for these individuals and the gain-loss utility stays the same. Taken together, increasing the fixed payment either increases individual effort or leaves effort unchanged:

> **Hypothesis 1:** *Increasing the fixed payment f weakly increases average labor supply.*

Since individuals below f are induced to work more and individuals above f to work less, we should observe clumping of stopping decisions exactly at the fixed payment f:

> **Hypothesis 2:** *The probability to stop when $we = f_{LO}$ is higher in the LO-treatment than in the HI-treatment; the probability to stop when $we = f_{HI}$ is higher in the HI-treatment than in the LO-treatment.*

In addition, since the gain-loss utility does not depend on the distance between we

and f, we should see no treatment difference below f_{LO} or above f_{HI}. The treatment effect should only be driven by subjects stopping between or at f_{LO} and f_{HI}:

> **Hypothesis 3:** *The probability to stop at a given earnings level below f_{LO} or above f_{HI} is the same for LO- and HI-treatment.*

There is one other utility function that also predicts that subjects work harder when the fixed payment f is increased and that does not include reference dependence: if $U(x, e) = u(x - c(e))$ and subjects are very risk averse, then the fixed payment can have an effect on the labor supply decision. With this integrated utility function, a subject doesn't want to exert a lot of effort on days when he doesn't make a lot of money (even if there is no direct connection between effort and money). Applied to our experimental game, this utility function becomes:

$$U = \frac{1}{2}u(f - c(e)) + \frac{1}{2}u(we - c(e))$$

Labor supply increases in f for this utility function for the following reasoning: as f rises, the cost of effort decreases in the state of the world when the subject gets f since he is in the flatter part of the concave utility function. In the state of the world when he gets we, an increase of f leaves him at the same place in the utility function. So the increase of f reduces only the cost of effort but does not reduce the value of effort that is generated only in the state when he gets we. Therefore, a higher f leads to a higher average effort.

There are a couple of reasons why one should be skeptical that this utility function would drive a potential treatment effect: First of all, it is hard to see why working is more painful on days when your income is low. Second, Rabin (2000) has shown that individuals have to be basically risk neutral for small gambles (e.g., $x < 100$ euros) if one wants to use the same utility function to also explain their behavior in large gambles. And with risk-neutrality, the integrated utility function predicts no treatment difference in labor supply. In any case, this utility function makes only the same prediction as a reference-dependent utility function for the treatment comparison of average labor supply (Hypothesis 1). It does not predict a higher probability to stop when accumulated earnings equal the fixed payment f (Hypothesis 2); and it predicts the treatment effect to be present over the whole range of accumulated earnings, whereas reference-dependent models predict no treatment difference for subjects that stopped at accumulated earnings levels below f_{LO} or above f_{HI} (Hypothesis 3).

3.4 Results

In this section we present results of the experiment. The main variable of interest is the accumulated earnings at which a subject decided to stop.

> **Result 1:** *Labor supply is significantly different between treatments. Subjects work more when the fixed payment is larger.*

In the LO-treatment with fixed payment $f = 3$ euros, subjects stop working after accumulating 7.37 euros on average. In the HI-treatment with $f = 7$ euros, subjects stop on average at 9.22 euros. The treatment difference of 1.85 euros is quite large, given that the treatment manipulation is only $7 - 3 = 4$ euros. The marginal effect is 25.1 percent. The difference between treatments is significant (Mann-Whitney U-test, $p = 0.015$)[5]. We obtain the same result if we compare the distribution of stopping decisions: A two-sample Kolmogorov-Smirnov test rejects the equality of distributions between treatments ($p = 0.005$). These results hold if we consider the time spent working as alternative measure of labor supply: subjects in the LO-treatment work on average 31.7 minutes, while subjects in the HI-treatment work on average 6.4 minutes longer, a marginal effect of 20.1 percent. This difference is statistically significant (U-test, $p = 0.034$; Kolmogorov-Smirnov test, $p = 0.085$).

The result is further confirmed by OLS regressions where we can control for other potential determinants of work effort (see Table 3.1). Without any controls, the treatment effect of 1.85 euros has a p-value of 0.046 (column 1). Adding controls for productivity[6], gender, dummies for time of day, and outside temperature (exper-

[5] All p-values in this chapter refer to two-sided tests.

[6] We use average time per answer in the first stage as an indicator for productivity in the main stage (measured in seconds times -1). A positive coefficient thus indicates that faster subjects complete more tables. This measure is not influenced by the treatment manipulation since, during the first stage, subjects did not know yet about the fixed payment and the exact procedure of the main stage; consequently, answering speed in the first stage is not significantly different between

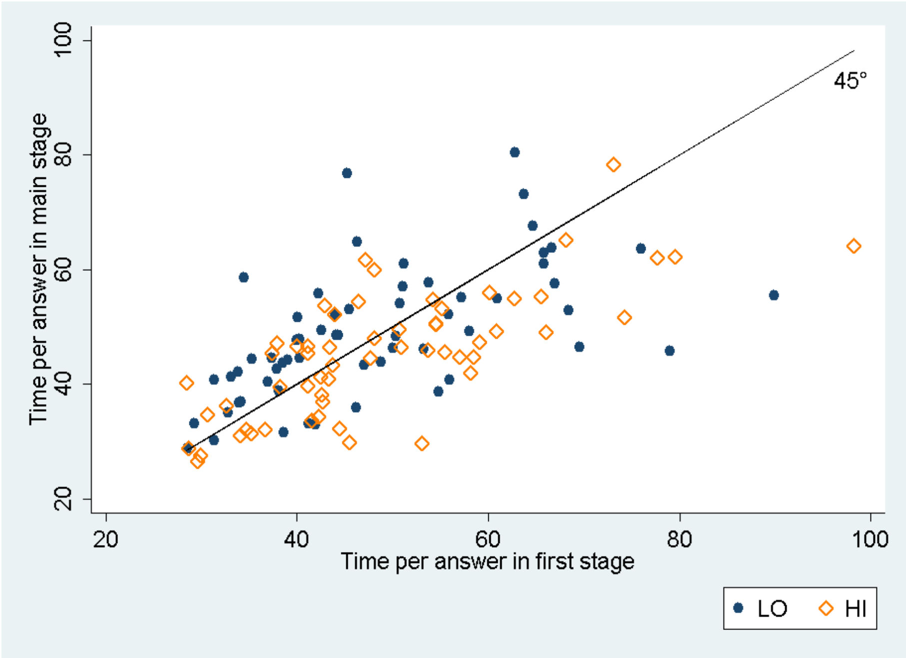


Figure 3.1: *Average time per answer during first and main stage.*

iments took place in summer) increases the treatment coefficient to 2.13 euros and reduces its p-value to 0.018 (columns 2 and 3). The only significant control variable is productivity. These results hold similarly if we take the time subjects spent working as dependent variable (see columns 1–3 in Table C.2 in Appendix C.6). Subjects can, however, only work between 0 and 60 minutes. We therefore also perform tobit regressions to account for this censoring (columns 1–3 in Table C.3). This does not alter results.[7]

As shown in Section 3.3, the model of Kőszegi & Rabin (2006) does not only predict treatments to be different but to be different in a very special way. First of all, they predict a higher probability to stop when the accumulated earnings equal the fixed payment. This is what we find in our data.

treatments (U-test, $p = 0.567$). Additionally, Figure 3.1 shows that the answering speed is very stable between the two stages. The Spearman rank correlation coefficient is 0.690 ($p < 0.001$). Using instead average time per table (i.e., per *correct* answer) or number of completed tables during the first stage does not change results.

[7]Censoring is not an issue if we take earnings as dependent variable; earnings are neither bounded above nor below (since subjects could make losses by miscounting tables thrice).

Table 3.1: Treatment Difference (Accumulated Earnings) in OLS Regressions

	Dependent Variable: Accumulated earnings (in euro) at which a subject stopped								
	Full sample			Stopped between 3 and 7 euros			Stopped below 3 or above 7 euros		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1 if HI-treatment	1.850**	2.006**	2.110**	1.564***	1.563***	1.360***	1.239	1.486	1.954
	(0.917)	(0.873)	(0.911)	(0.343)	(0.348)	(0.415)	(1.349)	(1.253)	(1.408)
Productivity		0.116***	0.109***		0.000	0.022		0.163***	0.137**
		(0.031)	(0.035)		(0.012)	(0.018)		(0.048)	(0.060)
1 if Female			0.140			0.663			-0.056
			(1.018)			(0.487)			(1.494)
Outside Temperature			-0.065			0.035			-0.147
(in °C)			(0.140)			(0.060)			(0.232)
Controls for time of day	No	No	Yes	No	No	Yes	No	No	Yes
Constant	7.370***	12.972***	12.713***	4.386***	4.386***	4.600***	10.161***	17.682***	16.155**
	(0.648)	(1.644)	(3.984)	(0.231)	(0.655)	(1.607)	(0.989)	(2.389)	(6.740)
N.Obs.	120	120	120	53	53	53	67	67	67
Prob > F	0.046	0.000	0.064	0.000	0.000	0.107	0.362	0.003	0.202
Adjusted R^2	0.025	0.119	0.093	0.275	0.261	0.194	-0.002	0.138	0.087

Table 3.1: Treatment Difference (Accumulated Earnings) in OLS Regressions. Notes: The dependent variable is the level of accumulated earnings (in euro) at which a subject stopped working. The proxy for productivity is the time subjects needed per answer during the first stage (in seconds times -1); a positive coefficient thus indicates that faster subjects complete more tables. Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Result 2: *The probability to stop when accumulated earnings are equal to the amount of the fixed payment is higher compared to the same earnings level in the other treatment. The modal choice in both treatments is to stop exactly when accumulated earnings equal the fixed payment.*

Figure 3.2 shows a histogram of accumulated earnings for each treatment (LO in the top panel, HI in the bottom panel). First of all, stopping decisions are dispersed over a wide range. Some subjects stop directly, others work for up to 25 euros. This is what one would expect given that productivity and cost of effort differ across subjects. But there are systematic differences between treatments: in the LO-treatment, many subjects stop at or around 3 euros; in the HI-treatment, almost nobody stops at or around 3 euros. By contrast, in the HI-treatment many subjects stop at or around 7 euros; in the LO-treatment very few subjects stop in this range. The modal choice in both treatments is to stop exactly when accumulated earnings equal the fixed payment. These treatment differences are highly significant. 15.0 percent of subjects in the LO-treatment stop exactly at 3 euros and only 1.7 percent in the HI-treatment (U-test, $p = 0.009$). Only 3.3 percent of subjects in the LO-treatment stop at 7 euros and 16.7 percent in the HI-treatment (U-test, $p = 0.015$). The same results obtain if we compare the number of subjects stopping in a range around 3 and 7 euros. Between 2 and 4 euros, 30.0 and 5.0 percent of subjects stop in the LO- and HI-treatment, respectively (U-test, $p < 0.001$); between 6 and 8 euros, these figures are 13.3 and 36.7 percent, respectively (U-test, $p = 0.003$).

We asked subjects in a questionnaire about reasons for their stopping decision. Answers were given in free form without suggestion of possible reasons. The free form answers were then classified in 14 categories, defined such that as few answers as possible were left unclassified. Answers could be classified under multiple categories. Of those subjects stopping exactly when accumulated earnings equal the fixed payment, 83.3 percent named reasons like that they wanted to "avoid uncertainty", that they feared "losing their earnings" when they worked for more than the fixed payment, or that they wanted to "make sure" to get the amount of the fixed payment by working at least that much. These answers suggest that loss aversion or disappointment aversion drive our results. Indeed, subjects mentioning these reasons (uncertainty, loss, ...) stop significantly closer to the fixed payment than

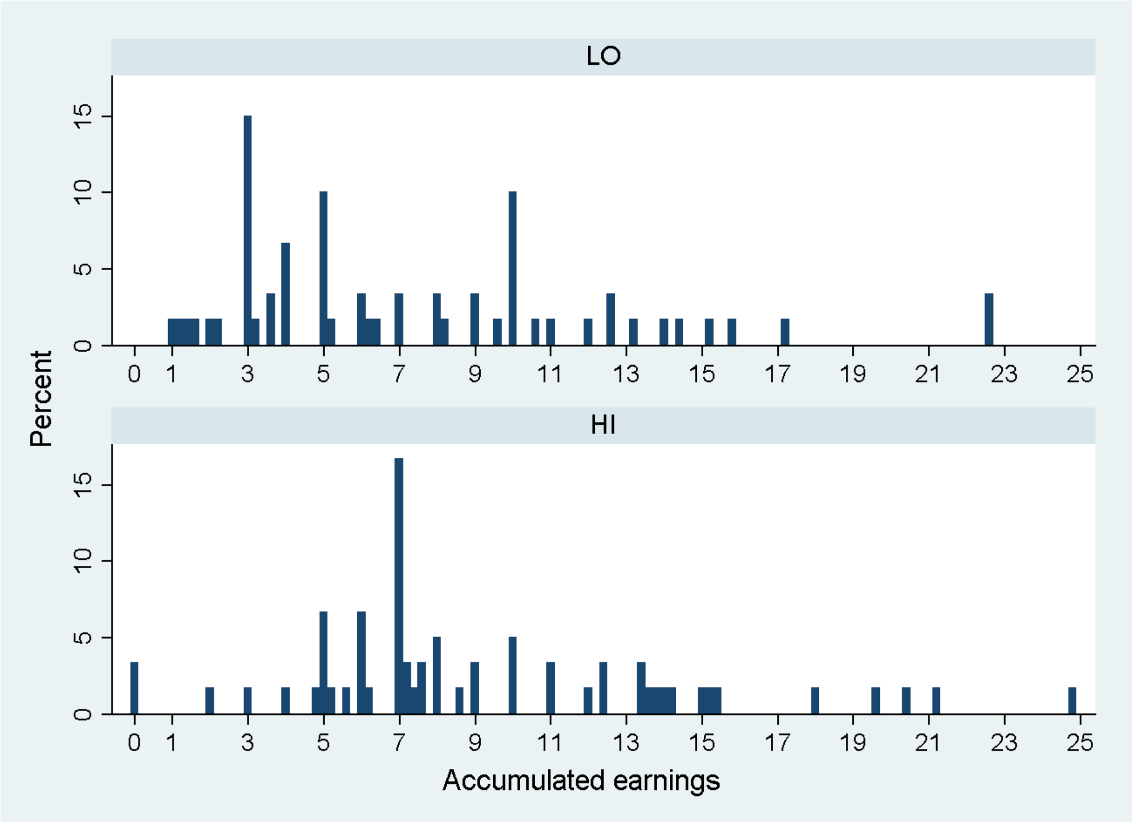


Figure 3.2: *Histogram of accumulated earnings (in euro) at which a subject stopped.*

other subjects ($\Delta = 1.52$ euros vs. $\Delta = 6.50$ euros; U-test, $p < 0.001$).[8]

Our setup also allows for tentative inferences about the functional form of reference dependence and loss aversion. If gain-loss utility is linear, the distance to the reference point does not appear in the first-order condition. Thus, once a subject has worked for more than 7 euros, the influence of reference dependence is the same in both treatments. We should therefore observe no treatment effect beyond 7 euros. The same reasoning applies to the range below 3 euros. Most theories of reference dependence indeed assume such a linear gain-loss utility. We find evidence in support of this assumption:

[8] Other common reasons to stop were "to maximize the possible monetary gain" (24.7 percent), especially by subjects who worked full time (77.8 percent of these subjects state this reason); many subjects mentioned to stop because they got tired or that they did more and more mistakes (22.1 percent).

Result 3: *The treatment difference is driven by subjects stopping in the range of 3 to 7 euros. For the group of subjects who stop below 3 euros or above 7 euros, treatments are not different.*

In Figure 3.3, the cumulative frequency of accumulated earnings is depicted. Between 3 and 7 euros, the two curves are clearly separated: subjects in the HI-treatment tend to stop at higher earnings levels. The two curves are much closer below 3 euros and beyond 7 euros. Indeed, the two distributions cross both below 3 euros and above 7 euros.

If we compare the two treatments only for subjects who stop in the range of 3 to 7 euros, we find a large and significant difference (U-test, $p < 0.001$). The group of all other subjects displays no treatment difference (U-test, $p = 0.571$). The latter result holds true if we consider only subjects who stopped before reaching 3 euros ($p = 0.241$) or only subjects who stopped after 7 euros ($p = 0.851$). These results are very robust. They hold if we test for treatment differences using a Kolmogorov-Smirnov test (between: $p = 0.003$; above/below: $p = 0.688$); we also obtain the same result in OLS regressions: for subjects stopping in the range of 3 to 7 euros, the p-value of the treatment dummy is smaller than 0.001 in a regression without controls. With the controls described above, the p-value is 0.003 (see columns 4 to 6 in Table 3.1). Treatments are *not* different for the remaining subjects (without controls: $p = 0.362$, with controls: $p = 0.172$, see columns 7 to 9 in Table 3.1). The size of the coefficient of the treatment dummy is not very different for the two groups, but keep in mind that the difference in the between group is bounded by 4 while it can be much larger in the above/below group. All these results also hold if we take time spent working as dependent variable.[9]

As pointed out in Section 3.3, subjects could work more in the HI-treatment also because of a non-reference-dependent utility function with the form $U = u(x - c(e))$

[9]Performing the same tests with time spent working as dependent variable yields the following results: U-test for subjects who stopped between 3 and 7 euros: $p = 0.029$; U-test above and below: $p = 0.213$; Kolmogorov-Smirnov test between: $p = 0.035$; Kolmogorov-Smirnov test above/below: $p = 0.266$. Columns 4 to 6 of Table C.2 in Appendix C.6 present OLS regressions for subjects who stopped between 3 and 7 euros; columns 7 to 9 show results for the remaining subjects. Corresponding tobit estimates (as subjects could only work between 0 and 60 minutes) are presented in columns 4 to 9 of Table C.3.

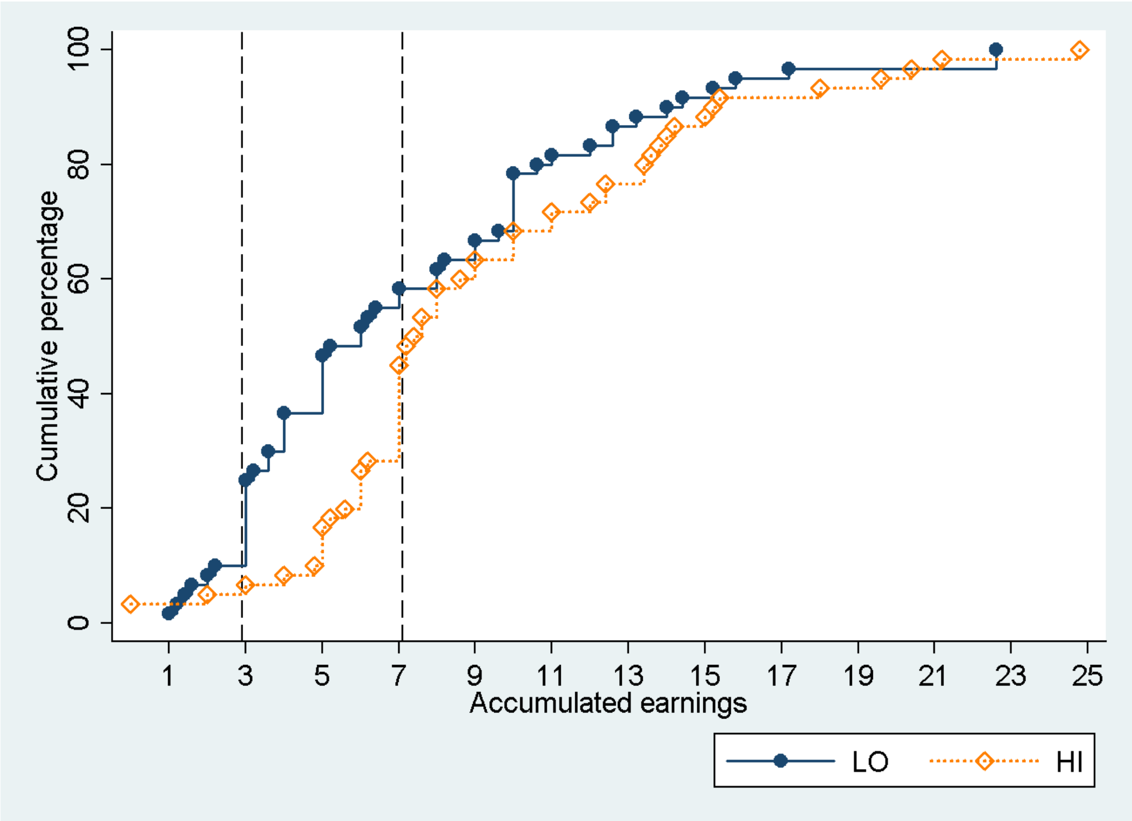


Figure 3.3: *Cumulative frequency of accumulated earnings (in euro) at which a subject stopped.*

where $u(\cdot)$ is very concave. The two previous results are already in contrast with such a utility function: it does not predict a higher propensity to stop when accumulated earnings equal the fixed payment (which we find however); and it predicts a treatment difference over the whole range of accumulated earnings (which we don't find). Additionally, the concavity needed to explain behavior for the small stakes in our experiment implies unreasonable behavior for large gambles: Rabin (2000) has shown that if a subject rejects risky gambles involving small stakes because of risk aversion (i.e., a concave utility function without a kink), the same subject will reject large gambles even with a very high (positive) expected value. Vice versa, if a subject accepts these large gambles, his rejecting small gambles cannot be explained by risk aversion. We therefore asked our subjects in a questionnaire whether they would accept a gamble in which they could lose 24 euros with 50 percent probability and win 10,000 euros with 50 percent probability. If subjects accept this gamble, their behavior in the experiment cannot be driven by risk aversion. Perhaps not surprisingly, 92.2 percent of subjects accept the gamble. If we exclude the remaining 7.8 percent, all results presented above continue to hold.

3.5 Conclusion

In a simple real-effort laboratory experiment, we tested theories of reference-dependent preferences that assume the reference point to be a function of individual expectations. Influences of risk aversion and status-quo reference dependence on the treatment effect were excluded by our experimental design. We find strong support for models of reference-dependent preferences and provide empirical evidence that theories like in Kőszegi & Rabin (2006) describe phenomena that cannot be caught with existing models of expected utility theory or status-quo loss aversion.

Our results are also relevant for the literature on reference points and labor supply. Studies in this literature use field data on worker effort choices, and have contributed to a lively debate regarding whether the response of effort to changes in incentives is consistent with the standard intertemporal substitution of labor and leisure or rather with loss aversion around a daily reference income (e.g., Camerer et al. 1997, Farber 2005, Fehr & Götte 2007).[10] In addition to the challenge of establishing exogenous variation in incentives, a key lacuna in this literature is that the reference point is unobserved. Our experiment makes the reference point known to the researcher, and also allows exogenous variation in the reference point. The experiment provides strong evidence that reference points can have a causal impact on labor supply. As noted by Kőszegi & Rabin (2007), if reference points are based on expectations, anticipated changes in incentives should not distort behavior relative to the standard theory, given that expectations adjust to reflect the anticipated change. For example, if an individual expects the hourly wage to be low on a given day, earning a small amount does not feel like a loss. But if the hourly wage is unexpectedly low, this does feel like a loss relative to expectation, and can induce workers to work even harder to try to reach their expectation, contrary to the standard prediction on intertemporal substitution which implies that workers should decrease effort when the wage is temporarily low. This distinction potentially helps reconcile some of the apparently conflicting findings in the field evidence, as in Camerer et al. (1997) and Farber (2005).

With the setup at hand, we cannot distinguish between different expectation-based models of reference-dependent preferences. Especially, we cannot test which method to map expectations into the reference point yields better predictions: assuming that the reference point is the mean of expectation (like in, e.g., Bell 1985) or assuming that the reference point is the whole distribution of expectations (like in, e.g., Kőszegi & Rabin 2006). Both of these assumptions predict a higher probability to stop when accumulated earnings equal the fixed payment. But a simple extension of our experimental design can distinguish between these models: if subjects' final payoff is determined by a lottery over *two* distinct fixed payments and accumulated earnings, predictions are different. Models like the one of Bell (1985) predict a higher probability to stop when accumulated earnings equal the mean of the two fixed payments but not when they equal the two fixed payments. Models like in Kőszegi & Rabin (2006) predict a higher probability to stop at the two fixed payments but not at the mean.

[10]For surveys of the literature see Götte et al. (2004) and Farber (2004).

Bibliography

Akerlof, G. (1982), 'Labor contracts as partial gift exchange', *Quarterly Journal of Economics* **97**, 543–569.

Akerlof, G. A. & Yellen, J. L. (1990), 'The fair wage-effort hypothesis and unemployment', *Quarterly Journal of Economics* **105**, 255–283.

Andreoni, J. & Miller, J. (2002), 'Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism', *Econometrica* **70**(2), 737–753.

Arkes, H. R., Joyner, C. A., Pezzo, M. V., Nash, J., Siegel-Jacobs, K. & Stone, E. (1994), 'The psychology of windfall gains', *Organizational Behavior and Human Decision Processes* **59**(3), 331–347.

Ashraf, N., Camerer, C. & Loewenstein, G. (2005), 'Adam Smith, Behavioral Economist', *Journal of Economic Perspectives* **19**(3), 131–145.

Babcock, L., Wang, X. & Loewenstein, G. (1996), 'Choosing the wrong pond: Social comparisons in negotiations that reflect a self-serving bias', *Quarterly Journal of Economics* **111**, 1–21.

Baker, G. P., Jensen, M. C. & Murphy, K. J. (1988), 'Compensation and incentives: Practice vs. theory', *Journal of Finance* **43**(3), 593–616.

Barberis, N. & Huang, M. (2001), 'Mental accounting, loss aversion, and individual stock returns', *Journal of Finance* **56**(4), 1247–1292.

Barberis, N., Huang, M. & Santos, T. (2001), 'Prospect theory and asset prices', *Quarterly Journal of Economics* **116**(1), 1–53.

Barberis, N., Huang, M. & Thaler, R. (2006), 'Individual preferences, monetary gambles, and stock market participation: A case for narrow framing', *American Economic Review* **96**(4), 1069–1090.

Bartling, B. & von Siemens, F. (2004), 'Inequity aversion and moral hazard with multiple agents', *University of Munich Working Paper* .

Bell, D. E. (1985), 'Disappointment in decision making under uncertainty', *Operations Research* **33**(1), 1–27.

Benartzi, S. & Thaler, R. H. (1995), 'Myopic loss aversion and the equity premium puzzle', *Quarterly Journal of Economics* **110**(1), 73–92.

Benjamin, D. J., Brown, S. A. & Shapiro, J. M. (2006), 'Who is 'behavioral'?', *Harvard University Working Paper* .

Bernheim, D. (1994), 'A theory of conformity', *Journal of Political Economy* **102**(5), 842–877.

Bewley, T. (1999), *Why Wages Don't Fall During a Recession*, Harvard University Press, Cambridge, MA.

Blow, L., Walker, I. & Zhu, Y. (2004), 'Who benefits from child benefit?', *Working Paper, University of Warwick* .

Bolton, G. & Ockenfels, A. (2000), 'Erc: A theory of equity, reciprocity, and competition', *American Economic Review* **90**, 166–193.

Bowles, S. & Park, Y. (2005), 'Emulation, inequality, and work hours: Was thorsten veblen right?', *The Economic Journal* **115**, F397–F412.

Camerer, C. (2002), *Behavioral Game Theory*, Princeton University Press, Princeton, NJ.

Camerer, C., Babcock, L., Loewenstein, G. & Thaler, R. (1997), 'Labor supply of New York City cabdrivers: One day at a time', *Quarterly Journal of Economics* **112**(2), 407–41.

Casari, M., Ham, J. C. & Kagel, J. H. (2007), 'Selection bias, demographic effects, and ability effects in common value auction experiments', *American Economic Review* **97**(4), 1278–1304.

Charness, G. & Kuhn, P. (2005), 'Pay inequality, pay secrecy, and effort: Theory and evidence', *NBER Working Paper* **11786**.

Cherry, T. L., Frykblom, P. & Shogren, J. F. (2002), 'Hardnose the dictator', *American Economic Review* **92**(4), 1218–1221.

Clark, A. E. & Oswald, A. J. (1996), 'Satisfaction and comparison income', *Journal of Public Economics* **61**(3), 359–381.

Clark, A., Masclet, D. & Villeval, M.-C. (2006), 'Effort and comparison income: Survey and experimental evidence', *GATE Working Paper 06-01* .

Demougin, D. & Fluet, C. (2003), 'Group vs. individual performance pay when workers are envious', *CIRPÉE Working Paper* .

Dufwenberg, M. & Kirchsteiger, G. (2004), 'A theory of sequential reciprocity', *Games and Economic Behavior* **47**(2), 268–298.

Easterlin, R. (1974), Does economic growth improve the human lot? Some empirical evidence, *in* P. David & M. Reder, eds, 'Nations and households in economic growth: Essays in honor of Moses Abramovitz', Academic Press, New York, pp. 89–125.

Easterlin, R. A. (2001), 'Income and happiness: Towards a unified theory', *The Economic Journal* **111**, 465–484.

Epley, N., Mak, D. & Idson, L. C. (2006), 'Bonus or Rebate?: The Impact of Income Framing on Spending and Saving', *Journal of Behavioral Decision Making* **19**, 213–227.

Erev, I., Bornstein, G. & Galili, R. (1993), 'Constructive intergroup competition as a solution to the free rider problem: A field experiment', *Journal of Experimental Social Psychology* **29**, 463–478.

Fack, G. (2006), 'Are housing benefit an effective way to redistribute income? Evidence from a natural experiment in France', *Labour Economics* **13**(6), 747–771.

Falk, A. & Fischbacher, U. (2006), 'A theory of reciprocity', *Games and Economic Behavior* **54**(2), 293–315.

Falk, A. & Ichino, A. (2006), 'Clean evidence on peer effects', *Journal of Labor Economics* **24**(1), 39–57.

Farber, H. S. (2004), 'Reference-dependent preferences and labor supply: The case of New York City cabdrivers', *Princeton University Working Paper 497* .

Farber, H. S. (2005), 'Is tomorrow another day? The labor supply of New York City cabdrivers', *Journal of Political Economy* **113**(1), 46–82.

Fehr, E. & Falk, A. (2002), 'Psychological foundations of incentives', *European Economic Review* **46**, 687–724.

Fehr, E. & Gächter, S. (1998), 'Reciprocity and Economics: The Economic Implications of Homo Reciprocans', *European Economic Review* **42**, 845–859.

Fehr, E. & Gächter, S. (2000), 'Fairness and Retaliation: The Economics of Reciprocity', *Journal of Economic Perspectives* **14**, 159–181.

Fehr, E., Gächter, S. & Kirchsteiger, G. (1997), 'Reciprocity as a contract enforcement device', *Econometrica* **65**, 833–860.

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

Fehr, E., Kirchsteiger, G. & Riedl, A. (1993), 'Does Fairness Prevent Market Clearing? An Experimental Investigation', *Quarterly Journal of Economics* **108**, 437–460.

Fehr, E., Klein, A. & Schmidt, K. M. (forthcoming), 'Fairness and contract design', *Econometrica* .

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

Fehr, E. & Schmidt, K. M. (2003), Theories of Fairness and Reciprocity - Evidence and Economic Applications, *in* M. Dewatripont, L. Hansen & S. Turnovsky, eds, 'Advances in Economics and Econometrics - 8th World Congress, Econometric Society Monographs', Cambridge University Press, Cambridge.

Fischbacher, U. (2007), 'z-Tree: Zurich Toolbox for Ready-made Economic Experiments', *Experimental Economics* **10**(2), 171–178.

Frederick, S. (2005), 'Cognitive reflection and decision making', *Journal of Economic Perspectives* **19**(4), 25–42.

Fudenberg, D. & Levine, D. (2006), 'A dual self model of impulse control', *American Economic Review* **96**, 1449–1476.

Genesove, D. & Mayer, C. (2001), 'Loss aversion and seller behavior: Evidence from the housing market', *Quarterly Journal of Economics* **116**(4), 1233–1260.

Gibbons, S. & Manning, A. (2006), 'The incidence of UK housing benefit: Evidence from the 1990s reforms', *Journal of Public Economics* **90**(4), 799–822.

Gneezy, U. (2006), 'Bonuses versus predetermined wages: A study of incentives and reciprocity', *mimeo* .

Gneezy, U. & Potters, J. (1997), 'An experiment on risk taking and evaluation periods', *Quarterly Journal of Economics* **112**(2), 631–645.

Götte, L., Fehr, E. & Huffman, D. (2004), 'Loss aversion and labor supply', *Journal of the European Economic Association* **2**(2), 216–228.

Gul, F. (1991), 'A theory of disappointment aversion', *Econometrica* **95**(3), 667–686.

Hannan, R. L., Kagel, J. H. & Moser, D. V. (2002), 'Partial Gift Exchange in an Experimental Labor Market: Impact of Subject Population Differences, Productivity Differences, and Effort Requests on Behaviour', *Journal of Labor Economics* **20**(4).

Harrison, G. W. & List, J. A. (2004), 'Field experiments', *Journal of Economic Literature* **42**, 1009–1055.

Heath, C. & Soll, J. (1996), 'Mental budgeting and consumer decisions', *Journal of Consumer Research* **23**(1), 40–52.

Heidhues, P. & Kőszegi, B. (forthcoming), 'Competition and price variation when consumers are loss averse', *American Economic Review* .

Holmstrom, B. (1982), 'Moral hazard in teams', *Bell Journal of Economics* **13**(2), 324–340.

Itoh, H. (2004), 'Moral hazard and other-regarding preferences', *Japanese Economic Review* **55**(1), 18–45.

Kahneman, D., Knetsch, J. & Thaler, R. (1990), 'Experimental tests of the endowment effect and the Coase theorem', *Journal of Political Economy* **98**, 1325–1348.

Kahneman, D. & Lovallo, D. (1993), 'Timid choices and bold forecasts: A cognitive perspective on risk taking', *Management Science* **39**(1), 17–31.

Kahneman, D. & Tversky, A. (1979), 'Prospect theory: An analysis of decision under risk', *Econometrica* **47**, 263–291.

Kőszegi, B. & Rabin, M. (2006), 'A model of reference-dependent preferences', *Quarterly Journal of Economics* **121**(4), 1133–1165.

Kőszegi, B. & Rabin, M. (2007), 'Reference-dependent risk attitudes', *American Economic Review* **97**(4), 1047–1073.

Knez, M. & Simester, D. (2001), 'Firm-wide incentives and mutual monitoring at continental airlines', *Journal of Labor Economics* **19**(4), 743–772.

Kooreman, P. (2000), 'The labeling effect of a child benefit system', *American Economic Review* **90**(3), 571–583.

Laferrère, A. & Le Blanc, D. (2004), 'How do housing allowances affect rents? An empirical analysis of the French case', *Journal of Housing Economics* **13**(1), 36–67.

Laibson, D. (1997), 'Golden eggs and hyperbolic discounting', *The Quarterly Journal of Economics* **112**(2), 443–77.

Layard, R. (2005), Rethinking public economics: The implications of rivalry and habit, *in* L. Bruni & P. L. Porta, eds, 'Economics and Happiness', Oxford University Press, pp. 147–170.

Lazear, E. P. (1989), 'Pay equality and industrial politics', *Journal of Political Economy* **97**(3), 561–80.

Levin, L. (1998), 'Are assets fungible? Testing the behavioral theory of life-cycle savings', *Journal of Economic Behavior & Organization* **36**, 59–83.

Loewenstein, G., Thompson, L. & Bazerman, M. (1989), 'Social utility and decision making in interpersonal contexts', *Journal of Personality and Social Psychology* **57**, 426–441.

Loomes, G. & Sugden, R. (1986), 'Disappointment and dynamic consistency in choice under uncertainty', *The Review of Economic Studies* **53**(2), 271–282.

Loomes, G. & Sugden, R. (1987), 'Testing for regret and disappointment in choice under uncertainty', *Economic Journal* **97**, 118–129.

Maciejovsky, B., Kirchler, E. & Schwarzenberger, H. (2001), 'Mental accounting and the impact of tax penalty and audit frequency on the declaration of income: An experimental analysis', *Humboldt-University of Berlin Discussion Paper* .

Manski, C. F. (2000), 'Economic analysis of social interactions', *Journal of Economic Perspectives* **14**(3), 115–136.

Maximiano, S., Sloof, R. & Sonnemans, J. (forthcoming), 'Gift exchange in a multi-worker firm', *Economic Journal* .

Medoff, J. L. & Abraham, K. G. (1980), 'Experience, performance, and earnings', *The Quarterly Journal of Economics* **95**(4), 703–36.

Milgrom, P. & Roberts, J. (1992), *Economics, Organization and Management*, 1 edn, Prentice Hall, Englewood Cliffs, NJ.

Milkman, K. L., Beshears, J., Rogers, T. & Bazerman, M. H. (2007), 'Mental accounting and small windfall gains: Evidence from an online grocer', *Harvard University Discussion Paper* .

Modigliani, F. & Brumberg, R. (1954), Utility analysis and the consumption function: An interpretation of cross-section data, *in* K. K. Kurihara, ed., 'Post-Keynesian Economics', Rutgers University Press, New Brunswick, NJ.

Moffitt, R. (1989), 'Estimating the value of an in-kind transfer: The case of food stamps', *Econometrica* **57**, 385–409.

Munro, A. (2005), 'The economics of the winter fuel allowance', *mimeo* .

Neumark, D. & Postlewaite, A. (1998), 'Relative income concerns and the rise in married women's employment', *Journal of Public Economics* **70**, 157–183.

O'Curry, S. (1997), 'Income source effects', *DePaul University Discussion Paper* .

Odean, T. (1998), 'Are investors reluctant to realize their losses?', *Journal of Finance* **53**(5), 1775–1798.

Prelec, D. & Loewenstein, G. (1998), 'The red and the black: Mental accounting of savings and debts', *Marketing Science* **17**(1), 4–28.

Prendergast, C. (1999), 'The provision of incentives in firms', *Journal of Economic Perspectives* **37**, 7–63.

Rabin, M. (1993), 'Incorporating fairness into game theory and economics', *American Economic Review* **83**(5), 1281–1302.

Rabin, M. (2000), 'Risk aversion and expected-utility theory: A calibration theorem', *Econometrica* **68**(5), 1281–1292.

Rabin, M. & Weizsäcker, G. (2007), 'Narrow bracketing and dominated choices', *IZA Discussion Paper* .

Read, D., Loewenstein, G. & Rabin, M. (1999*a*), 'Choice Bracketing', *Journal of Risk and Uncertainty* **19**(1), 171–197.

Read, D., Loewenstein, G. & Rabin, M. (1999*b*), 'Choice bracketing', *Journal of Risk and Uncertainty* **19:1–3**, 171–197.

Shefrin, H. & Thaler, R. (1988), 'The behavioral life-cycle hypothesis', *Economic Inquiry* **26**, 609–643.

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

Susin, S. (2002), 'Rent vouchers and the price of low-income housing', *Journal of Public Economics* **83**(1), 109–152.

Thaler, R. (1980), 'Toward a positive theory of consumer choice', *Journal of Economic Behavior & Organization* **1**, 39–60.

Thaler, R. (1985), 'Mental accounting and consumer choice', *Marketing Science* **4**, 199–214.

Thaler, R. (1999), 'Mental accounting matters', *Journal of Behavioral Decision Making* **12**, 183–206.

Thaler, R. H. & Johnson, E. J. (1990), 'Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice', *Management Science* **36**(6), 643–660.

Thaler, R. H. & Sunstein, C. R. (2003), 'Libertarian paternalism', *American Economic Review* **93**, 175–179.

Thaler, R., Tversky, A., Kahneman, D. & Schwartz, A. (1997), 'The effect of myopia and loss aversion on risk taking: An experimental test', *Quarterly Journal of Economics* **112**(2), 647–661.

Thöni, C. & Gächter, S. (2005), 'Social interaction in the workplace', *mimeo* .

Torgler, B., Schmidt, S. L. & Frey, B. S. (2006), 'Relative income position and performance: an empirical panel analysis', *IEW Working Paper* **268**.

Tversky, A. & Kahneman, D. (1981), 'The framing of decisions and the psychology of choice', *Science* **211**, 453–458.

White, R. J. (2006), 'Format matters in the mental accounting of funds: The case of gift cards and cash gifts', *University of Waterloo Discussion Paper* .

Appendices

A.1 Predictions for Inequality-averse Players

In this appendix we analyze the behavior of the players if they maximize an inequality aversion utility function of the following form (see Fehr & Schmidt 1999):

$$U_i(\pi) = \pi_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max[\pi_j - \pi_i, 0] - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max[\pi_i - \pi_j, 0]$$

Here π is the vector of monetary payoffs of the players, n is the number of the players in the reference group, α_i represents the "envy" of player i if his monetary payoff is smaller than that of player j, and $\beta_i \in [0, 1)$, $\beta_i \leq \alpha_i$ is i's feeling of "guilt" if he has a larger monetary payoff than j.

We assume that the inequality averse players compare their monetary payoff to the (expected) monetary payoff of both other players in their firm. We first assume that the parametrization given by Fehr & Schmidt (1999) holds for our experimental subjects (section A.1.1) and subsequently calculate the equilibria with a slightly different distribution of the fairness parameters α and β (section A.1.2).

A.1.1 Parametrization by Fehr and Schmidt

As we will see below, the crucial part of the parameter distribution in this section is the distribution of the guilt aversion parameter β. Using data from various experiments, Fehr and Schmidt derive the following distribution for β: 30% of the population show no guilt aversion at all ($\beta = 0$), 30% are slightly guilt averse ($\beta = 0.25$), and 40% show stronger guilt aversion ($\beta = 0.6$). The distribution of the envy aversion parameter α is: 30% with $\alpha = 0$, 30% $\alpha = 0.5$, 30% $\alpha = 1$, 10%

$\alpha = 4$. As Fehr & Schmidt (1999) we assume that α and β are perfectly correlated. This last assumption plays only a role in section A.1.2.

With this distribution we are able to solve the stage game of our experiment by backward induction. It is clear that for our game, a selfish principal ($\beta = 0$) will never pay a wage on the last stage as this would only reduce his utility. For the wage decision of the inequality averse principals note first that only the feeling of guilt may cause principals to pay a positive wage. After the agents have chosen their effort and before wages are paid the monetary payoff of the principal is always greater or equal than the monetary payoff of each agent with our choice of parameters.[11] The decision problem of an inequality averse principal is therefore as follows:

$$\begin{aligned}\max_{w_1,w_2} u_P(w_1, w_2) &= 10(e_1 + e_2) - w_1 - w_2 \\ &\quad -\frac{\beta}{2}\max\{10(e_1 + e_2) - w_1 - w_2 - (w_1 - c(e_1)), 0\} \\ &\quad -\frac{\beta}{2}\max\{10(e_1 + e_2) - w_1 - w_2 - (w_2 - c(e_2)), 0\}\end{aligned}$$

s.t.

$$0 \leq w_i \leq 100$$

where e_1, e_2 are the efforts provided by agent 1 and 2 respectively and w_1, w_2 are the wages the principal pays to agent 1 and 2. Recall that the only difference between treatments is the following additional constraint that a principal in the group wage treatment faces:

$$w = w_1 = w_2$$

Rearranging terms, the objective function of the principal looks as follows:

$$\max_{w_1,w_2} 10(e_1 + e_2)[1 - \beta] - \frac{1}{2}\beta[c(e_1) + c(e_2)] + w_1(\frac{3}{2}\beta - 1) + w_2(\frac{3}{2}\beta - 1)$$

Under the assumptions on β given above this immediately implies that even the most inequality averse principals (the 40% with $\beta = 0.6$) will never pay a positive wage. An inequality reducing increase of the wage for an arbitrary agent by 1 unit costs 1 but gives the principal only an extra utility of $\frac{3}{2}\beta < 1$. Thus, in our setup stronger guilt aversion would be necessary to make positive wage choices optimal

[11] The principal has strictly positive payoffs as he gets all goods produced, i.e. $10(e_1 + e_2)$, whereas the agents have weakly negative payoffs as they have to bear the cost of production.

($\beta \geq \frac{2}{3}$). This is true for both treatments. Thus, under this distribution of β all principals pay no wage.

Given this, the analysis of workers' behavior is trivial. As they expect to receive a wage of zero with certainty, both selfish and inequality averse workers choose the minimal effort $e = 1$ on the first stage in both treatments. The Fehr-Schmidt model thus yields the same predictions as the subgame perfect Nash equilibrium assuming rational and selfish players.

A.1.2 Alternative Parametrization

We now assume that 40% of the players are slightly more guilt averse than in the distribution of preferences given above, i.e. $\beta > \frac{2}{3}$ instead of $\beta = 0.6$ for 40% of the subjects. As we have seen before, this means that positive wage payments can now be optimal for some of the principals. We leave the remaining distribution of parameters unchanged.

Individual Wage Treatment

We first consider the individual wage treatment. From the objective function above it is clear that a principal with $\beta > \frac{2}{3}$ wants to equalize the monetary payoffs of all members of his firm. If the "ex interim" difference (before paying wages) between his payoff and the agents' payoff is large enough, it is optimal for him to pay the following wages to the agents.

$$w_i^* = \frac{10}{3}e_i + \frac{10}{3}e_j + \frac{2}{3}c(e_i) - \frac{1}{3}c(e_j)$$

However, in our stage game, the monetary ex interim payoff of the principal is always so high that he can equalize all monetary payoffs by paying appropriate wages to the agents. For the following analysis we can thus assume that the principals with $\beta > \frac{2}{3}$ pay the wages given above in the IWT.

On the first stage, a completely selfish agent (i.e., an agent with $\alpha = \beta = 0$) maximizes his expected monetary payoff. Because he can only hope to receive a wage if he is matched with one of the principals with high enough β (which occurs with probability 0.4), he faces the following objective function:

$$\begin{aligned} \max_{e_i} \quad & -c(e_i) + \frac{4}{10}\left[\frac{10}{3}e_i + \frac{10}{3}e_j + \frac{2}{3}c(e_i) - \frac{1}{3}c(e_j)\right] \\ = \quad & \frac{4}{3}e_i - \frac{11}{15}c(e_i) + \frac{4}{3}e_j - \frac{2}{15}c(e_j) \end{aligned}$$

Taking into account the cost function described in Table 1.1 it is optimal for a selfish agent to choose effort $e_i^* = 3$.

An agent with positive fairness parameters takes the distribution of principals and co-workers he may be matched with as given and maximizes his expected Fehr-Schmidt utility. If he takes the effort difference to his co-worker as given, he fears to meet a principal with $\beta < \frac{2}{3}$ because in this situation no wages are paid and payoffs may therefore differ. We will show in the following that it is thus optimal for a fair agent to choose the minimal effort. We proceed as follows: We first consider the two cases where the inequity averse agent provides (weakly) higher / lower effort than his co-worker separately. We then show that his optimal strategy does not differ for these cases.

Under the assumption that the other agent, j, chooses a higher or equal effort than himself, the objective function of a fair agent i looks as follows:

$$\begin{aligned} \max_{e_i} \quad & -c(e_i) + \frac{4}{10}\left[\frac{2}{3}c(e_i) - \frac{1}{3}c(e_j) + \frac{10}{3}e_i + \frac{10}{3}e_j\right] \\ & -\frac{3}{10}\alpha[10e_i + 10e_j + c(e_i)] - \frac{3}{10}\beta\left[-c(e_i) + c(e_j)\right] \\ & = e_i\left[\frac{4}{3} - 3\alpha\right] + c(e_i)\left[-\frac{11}{15} - \frac{3}{10}\alpha + \frac{3}{10}\beta\right] + e_j\left[\frac{4}{3} - 3\alpha\right] + c(e_j)\left[-\frac{2}{15} - \frac{3}{10}\beta\right] \end{aligned}$$

As $\alpha \geq 0.5$ and $\alpha \geq \beta$ for the agents who are not completely selfish, the optimal choice is the exertion of minimal effort, $e_i^* = 1$.

If we assume that the co-worker j chooses a lower effort than i, i experiences more envy and less regret, but the optimal choice remains the same:

$$\begin{aligned} \max_{e_i} \quad & -c(e_i) + \frac{4}{10}\left[\frac{10}{3}e_i + \frac{10}{3}e_j + \frac{2}{3}c(e_i) - \frac{1}{3}c(e_j)\right] \\ & -\frac{3}{10}\alpha[10e_i + 10e_j + c(e_i)] - \frac{3}{10}\alpha\left[-c(e_j) + c(e_i)\right] \\ & = e_i\left[\frac{4}{3} - 3\alpha\right] + c(e_i)\left[-\frac{11}{15} - \frac{3}{5}\alpha\right] + e_j\left[\frac{4}{3} - 3\alpha\right] + c(e_j)\left[-\frac{2}{15} + \frac{3}{10}\beta\right] \end{aligned}$$

Minimal effort, $e_i^* = 1$, is again optimal for the fair player i.

We conclude that there is a unique equilibrium for the IWT, in which completely selfish agents choose $e = 3$, inequality averse agents choose $e = 1$, and only principals with $\beta > \frac{2}{3}$ pay positive wages such that ex post the monetary payoffs of all players are equal, i.e. $w_i = \frac{2}{3}c(e_i) - \frac{1}{3}c(e_j) + \frac{10}{3}e_i + \frac{10}{3}e_j$. The resulting average effort in the IWT is then 1.6 whereas the average wage is 8.69.

Equal Wage Treatment

In the equal wage treatment, the behavior of a highly guilt averse principal may differ compared to the individual wage treatment. If the two agents in his firm choose different effort levels, he can no longer equalize all payoffs. As he suffers more from disadvantageous inequality than from advantageous inequality ($\alpha \geq \beta$), he chooses the wage in order to match his monetary payoff with the "lazier" agent (the one with the lower effort level and thus higher payoff), thereby ensuring himself not to feel envious against that agent. The optimal wage payment w^* is therefore

$$w^* = \frac{10}{3}e_1 + \frac{10}{3}e_2 + \frac{1}{3}c(\min[e_1, e_2])$$

For the behavior of an agent it is again convenient to distinguish the case where he provides weakly lower effort than his co-worker from the case where his effort is higher. It will again be shown that all types of agents have a dominant strategy, so that finally this case distinction does not matter. In the case where the agent expects to exert lower or equal effort than his co-worker, his objective function looks as follows:

$$\begin{aligned}
\max_{e_i} \quad & -c(e_i) + \frac{4}{10}\left[\frac{10}{3}e_i + \frac{10}{3}e_j + \frac{1}{3}c(e_i)\right] - \frac{3}{10}\alpha\left[10e_i + 10e_j + c(e_i)\right] \\
& -\frac{3}{10}\beta\left[c(e_j) - c(e_i)\right] \\
& = e_i\left[\frac{4}{3} - 3\alpha\right] + c(e_i)\left[-\frac{13}{15} - \frac{3}{10}\alpha + \frac{3}{10}\beta\right] + e_j\left[\frac{4}{3} - 3\alpha\right] - \frac{3}{10}\beta c(e_j)
\end{aligned}$$

An agent who expects to exert higher effort than his co-worker, maximizes

$$\begin{aligned}
\max_{e_i} \quad & -c(e_i) + \frac{4}{10}\left[\frac{10}{3}e_i + \frac{10}{3}e_j + \frac{1}{3}c(e_j)\right] - \frac{3}{10}\alpha\left[10e_i + 10e_j + c(e_i)\right] \\
& -\frac{3}{10}\alpha\left[c(e_i) - c(e_j)\right] \\
& = e_i\left[\frac{4}{3} - 3\alpha\right] + c(e_i)\left[-1 - \frac{3}{5}\alpha\right] + e_j\left[\frac{4}{3} - 3\alpha\right] + c(e_j)\left[\frac{2}{15} + \frac{3}{10}\alpha\right]
\end{aligned}$$

It can be easily seen that agents with positive α and β (where $\alpha \geq \beta$) have a dominant strategy to exert minimal effort, $e_i^* = 1$ whereas completely selfish agents ($\alpha = \beta = 0$) choose $e_i^* = 3$ irrespective of what their co-worker does.

To summarize, the following strategies describe the unique equilibrium for the equal wage treatment: Selfish agents choose $e = 3$, fair agents choose $e = 1$, only principals with $\beta > \frac{2}{3}$ pay positive wages and match their payoff with the agent who exerted lower effort by paying $w = \frac{10}{3}e_1 + \frac{10}{3}e_2 + \frac{1}{3}c(\min[e_1, e_2])$. The resulting average effort for the EWT is thus again 1.6 whereas the average predicted wage is 8.58, marginally lower than the 8.69 in the IWT.

B.2 Instructions for Chapter 2

Welcome to today's decision experiment.

To start, please read these instructions carefully. At the end of the instructions you will find some example questions. The experiment starts as soon as all participants have answered these questions correctly.

Please note that it is not allowed to communicate with other participants of the experiment from now on. If this should happen, the experiment loses its scientific value and we have to stop the experiment. If you have any questions, please hold your hand out of the cubicle; we will then come to you.

The experiment consists of two parts. They will be called **work phase** and **shopping phase**. During the work phase you have the possibility to earn talers. You can then use these talers for shopping during the shopping phase. The value your purchases have for you will be denoted in points during the experiment. Directly after the experiment, the points you achieved will be summed up and paid in cash to you according to an exchange rate of

1 point = 0.01 euro

In addition, you receive **2.50 euro** for having showed up on time. The 2.50 euro will be paid after the experiment independently of your decisions and **additionally** to the amount you earn during the experiment.

Work phase

During the **work phase** you have the opportunity to earn 100 talers. The work consists of counting the number of zeros in tables filled with zeros and ones. Below, you see an example table with 3 rows and 8 columns. The tables used in the experiment are larger, they contain 10 rows and 30 columns.

Example of work phase

1	**1**	**1**	**0**	**0**	**0**	**1**	**0**
1	**0**	**1**	**0**	**1**	**1**	**0**	**1**
1	**0**	**0**	**0**	**1**	**0**	**1**	**1**

You earn the 100 talers if you succeed in finding the correct number of zeros in four tables within 15 minutes. If you do not succeed in finding the correct number of zeros in four tables you earn 10 talers instead.

Work phase screen

During the work phase, you will receive eight sheets with zeros and ones. Please begin on sheet 1 and count the number of zeros on this sheet. Enter the number of zeros in the input box in the middle of the computer screen. After entering the number click on the OK-button. If you entered the correct number, you may continue with sheet 2. If you entered a number that is higher by 1 or lower by 1 than the correct number, your number will also be rated as correct. If you enter a number that deviates by more than plus/minus 1 from the correct number, your input will be rated as false. You then have another two tries to enter the correct number for this sheet. Thus, you have three tries in total for each sheet. In the top-right hand corner of the screen, you can see the remaining time in seconds. The time starts at 900 seconds = 15 minutes and counts backwards.

Please note: the **red number** above the OK-button indicates the number of the current sheet. If you enter three times a wrong number for a sheet, the counter for the current sheet changes to the next sheet. If this occurs, please put the current sheet aside and start the next one.

You have a total of eight sheets at your disposal. As soon as you found the correct number of zeros on four sheets, the task is completed successfully and you receive 100 talers. **You then have finished the work phase.** If you do not succeed in completing the task within 15 minutes, you earn 10 talers instead.

Please note: Experience shows that is helpful to mark the 50th, 100th... counted zero. If you miscount in this case you do not have to start all over again but you can continue from the last marked zero.

Shopping phase

The **shopping phase** starts as soon as it has been determined for every participant if he or she completed the task of the work phase successfully. You will make **two** shopping decisions. Your credit balance is split equally between the two decisions. If you completed the task of the work phase successfully you have 100/2 = 50 talers at your disposal per purchasing decision, otherwise you have 10/2 = 5 talers.

During the shopping phase you can spend your money on two things that will be called **housing** and **clothing**. You decide which amount of housing and clothing you want to buy. Expenses for housing denote the rent of the apartment.

The value housing and clothing have for you are expressed in points that are exchanged into euro at the end of the experiment and paid out to you. How valuable a specific amount of housing or clothing is for you is denoted in two tables during the experiment. Below you see an example. In this example numbers of points and prices take on **different values** than in the experiment. The sole purpose of this example is to help you become familiar with the procedure of the purchasing decision.

Example of shopping phase

Housing	
Units	Points
0	0
1	6
2	11
3	15
4	18
5	20

Clothing	
Units	Points
0	0
1	16
2	24
3	27
4	29
5	30

Your credit balance
20 talers

Prices per unit
Housing: 4 talers
Clothing: 3 talers

In the left column of each table, the different amounts that are offered for sale are presented. The right column indicates how many points you get for the purchase of the corresponding amount. You can read from the table "Housing" that in this example 0 units of housing have a value of 0 points for you, 1 unit of housing has a value of 6 points, 2 units 11 points, and so on.

Your credit balance for the purchase is indicated in the top-right panel; in this example 20 talers. In the bottom-right panel you find the prices (in talers) for housing and clothing; prices are per unit. The **prices** for housing and clothing are **different**. The table "Prices per unit" shows that in this example a unit of housing costs 4 talers while clothing costs 3 talers per unit.

In the purchasing decision, you decide how many units of housing and how many units of clothing you want to buy. You can choose freely how many units to buy as long as the total price does not exceed your credit balance.

The **total price of your purchase** is calculated as follows:

$$\boxed{\begin{aligned}\text{Total price of purchase} = \quad & (\text{units of housing} \times \text{price per unit of housing}) \\ & + (\text{units of clothing} \times \text{price per unit of clothing})\end{aligned}}$$

As soon as you have decided how many units of housing and how many units of clothing to buy, it is determined how many points you will get for this decision. If you do not spend your entire credit balance, **the talers not spent are forfeited**. Additionally, talers from the first purchasing decision cannot be kept for the second purchasing decision.

The **total number of points** is calculated as follows:

Total number of points =	points for purchased units of housing
	+ points for purchased units of clothing

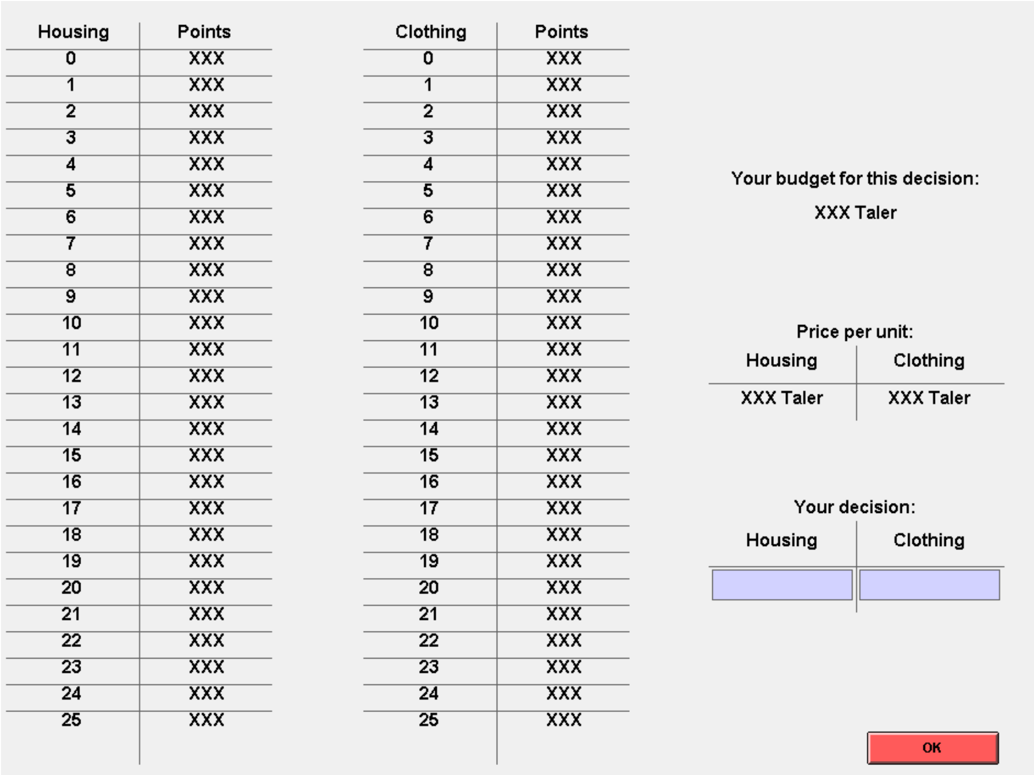
Example of a purchase

In the example mentioned above, you have a credit balance of 20 talers. Imagine you wanted to buy 3 units of housing and 2 units of clothing. Then you have to pay [(3 × price per unit of housing) + (2 × price per unit of clothing)] talers, i.e., 12+6 = 18 talers. This purchase is possible with your credit balance.
In the tables, you find the number of points you get for this purchase. You get **15 points** for 3 units of housing and **24 points** for 2 units of clothing. Your purchase would thus earn you $15 + 24 = 39$ **points**

Please note: It is only possible to buy **one** amount of each good. For example, if you want to buy altogether 4 units of clothing, the point value that is noted next to the number 4 (29 points) matters for you. You cannot buy first one unit of clothing and then another 3 units of clothing, for example.

On the computer, you make your decisions on the input screen of the shopping phase. Below you see a screen shot of this input screen. The screen contains all information that you need for your decision: tables for the point values of housing and clothing, your credit balance and the prices per unit. The actual point values and prices used in the experiment have been replaced with “XXX”.

Shopping phase screen



In the bottom-right hand corner of the screen, you can see two input fields. After having decided how many units of housing and of clothing to buy you enter your decision in these two fields and confirm your choice by clicking on the OK-button. **After having clicked on the OK-button you cannot change your decision anymore.** Your decision will be shown again on the screen. Please write your decision on the decision sheet that was handed out with these instructions. If you click on the OK-button although you would spend more talers than you have at your disposal, an error message is displayed and you have the possibility to correct your decision.

If you have any questions please hold your hand out of the cubicle; we will then come to you.

When all participants have answered the example questions correctly, the experiment starts with the working phase. When all participants have finished the working phase, you will be presented again short instructions for the first purchasing decision on the computer screen. Also for the second purchasing decision, the screen will show short instructions. As soon as all participants have taken the second purchasing decision the computer screen shows a questionnaire. After the questionnaire, the experiment is over.

Please answer the example questions handed out with these instructions before the experiment starts.

On-screen Instructions

Before the Working Phase

The working phase is about to start now. If you succeed in counting the correct number of zeros on four sheets within 15 minutes, you have completed the task successfully and you get 100 talers. If you do not succeed in completing the task successfully you get 10 talers instead.

Please click on the OK-button to start the working phase.

Before the First Purchasing Decision

You completed the task successfully. Your credit balance per purchasing decision is thus 50 talers.

In the following shopping phase you will make **two** purchasing decisions.

You decide how many units of housing and how many units of clothing to buy. You can read from the tables on the screen how many points you will get for your decision. If you do not spend all your credit balance, the talers not spent will be forfeited.

Before the Second Purchasing Decision

Lab Label treatment

For the second purchasing decision, you get a **housing subsidy** of **30 talers** in addition to your credit balance of 50 talers. You can spend the housing subsidy **only on housing**.

If the amount you spend on housing is lower than the amount of the housing subsidy, i.e., lower than 30 talers, the part of the subsidy that is not spent is **forfeited**.

The **housing subsidy** is the **only difference** compared to the first purchasing decision. All prices and point values remain the same.

Please note: When entering your purchasing decision, please report the **total** number of units you buy, no matter whether you paid them out of your **own credit balance** or out of the **housing subsidy**.

Lab Cash treatment

For the second purchasing decision, you get a **subsidy** of **30 talers** in addition to your credit balance of 50 talers. You can spend the subsidy on housing, on clothing or on both.

If you do not spend the whole subsidy, the part of it that is not spent is **forfeited**.

The **subsidy** is the **only difference** compared to the first purchasing decision. All prices and point values remain the same.

Please note: When entering your purchase decision, please report the **total** number of units you buy, no matter whether you paid them out of your **own credit balance** or out of the **subsidy**.

B.3 Moral Obligation Vignettes

Student-Support Scenario

> Mr Smith is a first-year Biology student who wants to apply for Bafög.[12] When he reads up on Bafög he notices that he has to specify the income of his parents and additionally his own wealth. He recently received part of his bequest, amounting to 32 000 euro. If he declares this amount his application will be rejected. He decides to not declare the bequest in his application in order to receive Bafög anyway.

What do you think about the behavior of Mr Smith on a scale from 1 to 6, where 1 means "Not appropriate at all" and 6 "Completely appropriate"?

Unemployment-Benefit Scenario

> Ms Newman has finished her studies of Law and is looking for a job. She has already found one but this position is only available in three months. She knows that she is eligible for unemployment benefit. She could easily bridge the time until the job starts since she has savings of 10 000 euro. Additionally, her parents support her with 800 euro per month until the new job starts. Ms Newman decides to claim unemployment benefit in addition, amounting to 300 euro per month.

What do you think about the behavior of Ms Newman on a scale from 1 to 6, where 1 means "Not appropriate at all" and 6 "Completely appropriate"?

[12] "Bafög" is the student support provided by the state in Germany. The amount depends on own income, own wealth and parents' income.

C.4 Instructions for Chapter 3

Below are the instructions of the HI-treatment translated into English. The only difference in the LO-treatment is that every occurrence of "7 euros" is replaced by "3 euros".

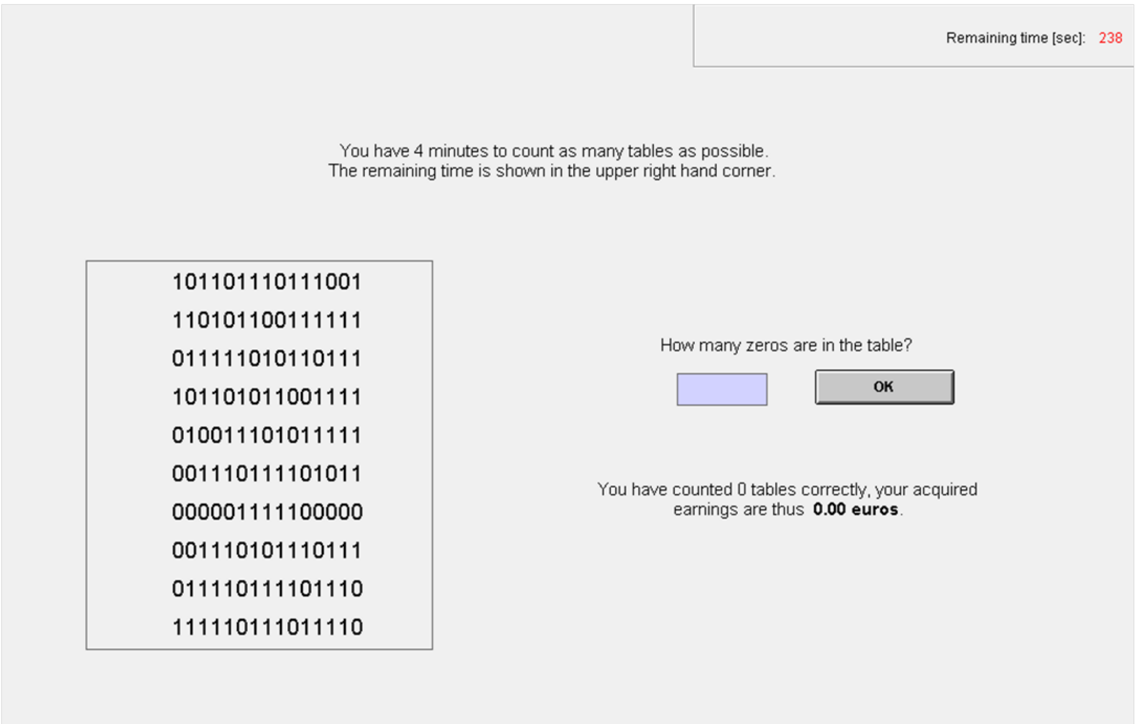
The experiment consists of two parts. Please start by reading the explanations for the first part carefully. You will receive the instructions for the second part of the experiment after the first part is finished.

For your arrival on time, you receive **5 euros** that will be paid to you at the end of the experiment. If you have any questions during the experiment please ask the experimenter. If you use the computer in an improper way you will be excluded from the experiment and from any payment.

Instructions for the first part of the experiment

What do you have to do?

In this part of the experiment your task is to count zeros in a series of tables. The figure shows the work screen you will use later:



Enter the number of zeros into the box on the right side of the screen. After you have entered the number, click the OK-button. If you enter the correct result, a new table will be generated. If your input was wrong, you have two additional tries to enter the correct number into the table. You therefore have a total of three tries to solve each table.

If you entered the correct number of zeroes you earn money: **You receive 10 cents for each table you solved correctly.**

If you enter three times a wrong number for a table, 10 cents will be subtracted from your earnings and a new table will then be generated. The earnings of this part of the experiment will be paid to you at the end of the experiment.

Example: You solve three tables correctly; you miscount one table once. You miscount a fourth table three times. Your earnings are therefore:

3 x 10c for the correctly counted tables

- 1 x 10c for the fourth table, which you miscounted three times.

thus a total of 20c.

You have 4 minutes until the first part of the experiment is over. The remaining time is displayed in the upper right hand corner of the screen.

Counting tips: Of course you can count the zeros any way you want. Speaking from experience, however, it is helpful to always count two zeros at once and multiply the resulting number by two at the end. In addition you miscount less frequently if you track the number you are currently counting with the mouse cursor.

Example question

Please answer the following question:

Assume you have solved 5 tables correctly, and miscounted two tables three times.

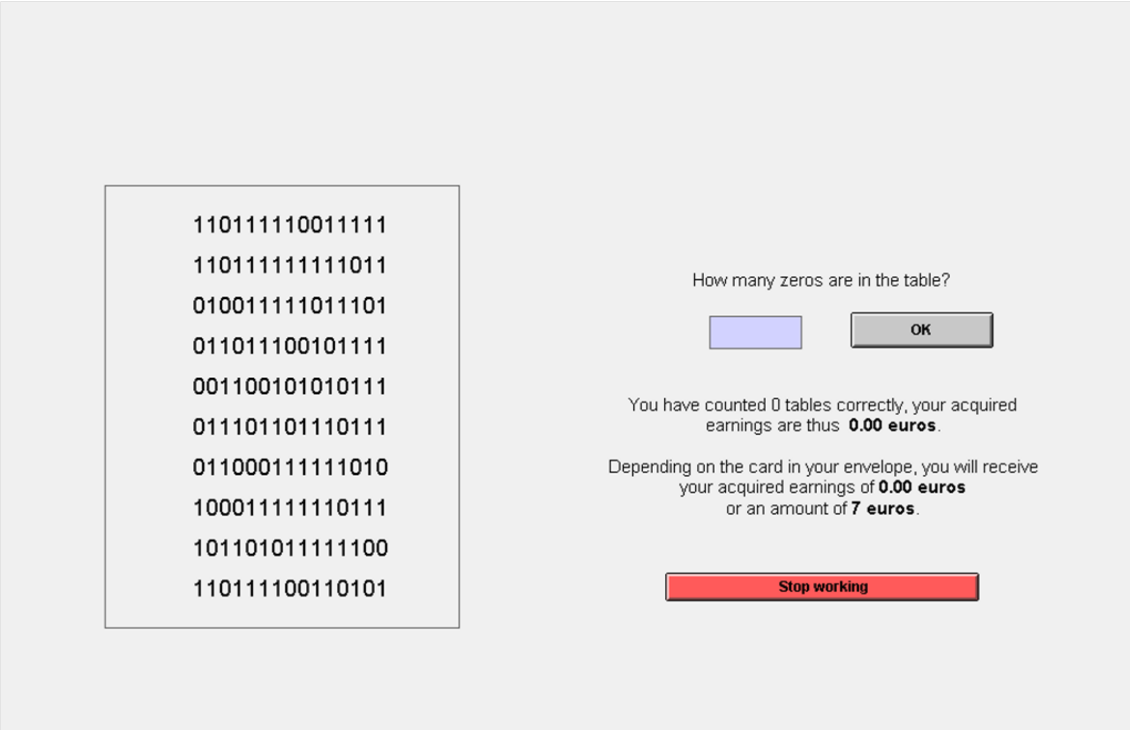
What are your acquired earnings? _____ euros

After you have answered the example question correctly, the experimenter will start the first part of the experiment.

Instructions for the second part of the experiment

What do you have to do?

The task in this part of the experiment is once again to count zeros in a series of tables. The figure shows the work screen you will use later:



New rules are now in effect, which did not apply in the first part:

- **For each correctly solved table you will be credited 20 cents.** After three wrong inputs 20 cents will be subtracted from your earnings.
- **It is possible to lose the acquired earnings from this part of the experiment:** there are two envelopes in front of you. One envelope contains a card with the text "acquired earnings", the other contains a card "7 euros". **You do not know which card is in which envelope.** Please choose one of the envelopes now and sign on the envelope.
- While you are working, the envelopes will remain in your room. After you have finished your task, we will open the envelopes. **You are not allowed to open** the envelopes before you have finished your task and one of the experimenters is with you.

- **If you have drawn the envelope with the card "acquired earnings", you will get your acquired earnings and not the 7 euros.**
- **If you have drawn the envelope with the card "7 euros", you will get 7 euros and not the acquired earnings.** The amount of 7 euros does not change, no matter how many tables you solved.
- After your work is done we will also open the envelope which you did not choose, so that you can check that the envelopes contained different cards.

Important: In this part of the experiment you can count zeros as long as you want. This means you can decide yourself when you want to stop working. You can work, however, at most 60 minutes.

If you want to stop counting, please click on the **red button** "stop working" and contact us by briefly stepping into the corridor. You will be paid in your room.

Example: You stop after ten minutes and have solved 24 tables correctly with no miscounts. Your acquired earnings are therefore $24 \times 20c = 4.80$ euros. The envelope chosen by you contains the card "acquired earnings". You therefore get 4.80 euros.

Example: You stop after 10 minutes and have solved 24 tables correctly with no miscounts. Your acquired earnings are therefore 4.80 euros. The envelope chosen by you contains the card "7 euros". You therefore get 7 euros instead of the 4.80 euros.

Example: You stop after 30 minutes and have solved 4 tables correctly and miscounted three times at a 5th table. Your acquired earnings are therefore $4 \times 20c - 1 \times 20c = 60c$. The envelope chosen by you contains the card "7 euros". You therefore get the amount of 7 euros instead of your acquired earnings of 60c.

Example questions

Please answer the following questions:

Assume you have solved 28 tables correctly within 20 minutes.

- What are your acquired earnings? _____euros
- How much money do you get if the envelope chosen by you contains the card "acquired earnings"? _____euros
- How much money do you get if the envelope chosen by you contains the card "7 euros"? _____euros

Assume you have solved 7 tables correctly within 15 minutes.

- What are your acquired earnings? _____euros
- How much money do you get if the envelope chosen by you contains the card "acquired earnings"? _____euros
- How much money do you get if the envelope chosen by you contains the card "7 euros"? _____euros

After you have answered the example questions correctly, the experimenter will start the second part of the experiment.

C.5 Photos of Experimental Rooms

C.6 Regressions of Total Time Worked

Table C.2: Treatment Difference (Time Worked) in OLS Regressions

Dependent Variable: Time spent working (in minutes) until a subject stopped

	Full sample			Stopped between 3 and 7 euros			Stopped below 3 or above 7 euros		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1 if HI-treatment	6.430**	6.525**	6.603*	5.307**	4.802**	4.482**	4.410	4.613	5.663
	(3.163)	(3.175)	(3.344)	(2.193)	(1.980)	(2.088)	(4.407)	(4.428)	(4.983)
Productivity		0.071	0.051		-0.246***	-0.176*		0.133	0.035
		(0.114)	(0.128)		(0.069)	(0.090)		(0.169)	(0.212)
1 if Female			1.962			0.515			3.619
			(3.735)			(2.446)			(5.287)
Outside Temperature			-0.476			-0.086			-0.588
(in °C)			(0.515)			(0.301)			(0.821)
Controls for time of day	No	No	Yes	No	No	Yes	No	No	Yes
Constant	31.715***	35.148***	38.565***	21.418***	8.912**	15.064*	41.347***	47.517***	44.323*
	(2.237)	(5.982)	(14.621)	(1.476)	(3.729)	(8.082)	(3.231)	(8.445)	(23.854)
N.Obs.	120	120	120	53	53	53	67	67	67
Prob > F	0.044	0.110	0.650	0.019	0.000	0.010	0.321	0.449	0.729
Adjusted R^2	0.03	0.02	-0.03	0.09	0.26	0.37	0.00	-0.01	-0.07

Table C.2: Treatment Difference (Time Worked) in OLS Regressions. Notes: The dependent variable is the time spent working (in minutes) until a subject stopped. The proxy for productivity is the time subjects needed per answer during the first stage (in seconds times -1); a positive coefficient thus indicates that faster subjects work longer. Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Table C.3: Treatment Difference (Time Worked) in Tobit Regressions

	Dependent Variable: Time spent working (in minutes) until a subject stopped								
	Full sample			Stopped between 3 and 7 euros			Stopped below 3 or above 7 euros		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1 if HI-treatment	7.927**	8.046**	8.234**	5.307**	4.802**	4.482***	7.735	7.894	9.366
	(3.841)	(3.840)	(3.728)	(2.151)	(1.923)	(1.622)	(6.404)	(6.396)	(6.274)
Productivity		0.080	0.053		-0.246***	-0.176**		0.092	0.017
		(0.138)	(0.142)		(0.067)	(0.070)		(0.244)	(0.264)
1 if Female			2.059			0.515			4.640
			(4.155)			(1.901)			(6.674)
Outside Temperature			-0.606			-0.086			-0.850
(in °C)			(0.567)			(0.234)			(0.998)
Controls for time of day	No	No	Yes	No	No	Yes	No	No	Yes
Constant	33.004***	36.874***	40.607**	21.418***	8.912**	15.064**	45.292***	49.515***	47.224
	(2.697)	(7.208)	(16.071)	(1.448)	(3.622)	(6.280)	(4.645)	(12.128)	(28.698)
N.Obs.	120	120	120	53	53	53	67	67	67
Prob > χ^2	0.041	0.104	0.589	0.017	0.000	0.000	0.231	0.452	0.437
Pseudo R^2	0.01	0.01	0.02	0.02	0.05	0.13	0.00	0.00	0.04

Table C.3: Treatment Difference (Time Worked) in Tobit Regressions. Notes: The dependent variable is the time spent working (in minutes) until a subject stopped. The lower and upper limits are 0 and 60 minutes. The proxy for productivity is the time subjects needed per answer during the first stage (in seconds times -1); a positive coefficient thus indicates that faster subjects worker longer. Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.