Fungibility, Labels, and Consumption[†]

Johannes Abeler[*] **Felix Marklein**[*]

University of Bonn and IZA Bonn

February 22, 2007

The fungibility of money is a central principle in economics. It means that any unit of money is substitutable for another. For consumption decisions, fungibility implies the specific composition of income to be irrelevant. In this paper, we show experimentally that even in a simple, incentivized setup many subjects do not treat money as fungible. Subjects are influenced by a label attached to a part of their budget and they change consumption in line with the suggestion of the label. Subjects with lower mathematical ability are more likely to exhibit this effect. An additional vignette survey confirms the experimental findings. Our results lend support to behavioral models like narrow bracketing or mental accounting and suggest, amongst others, that in-kind benefits distort consumption more than usually assumed.

JEL classification: C91, D01, H31, I38.

Keywords: Fungibility, In-Kind Benefits, Mental Accounting, Inframarginal Consumers, Laboratory Experiment

---

[†]Financial support from the DFG (through GRK 629) and from the Bonn Graduate School of Economics is gratefully acknowledged. We thank Steffen Altmann, Stefano DellaVigna, Kai Ebenrett, Gabrielle Fack, Simon Gächter, Uri Gneezy, David Huffman, Winfried Koeniger, Peter Kooreman, Anne Laferrère, Ulrike Malmendier, Daniel Hamermesh, Bentley MacLeod, Wolfram Merzyn, Susanne Ohlendorf, Robert Oxoby, Burkhard Schipper, Ian Walker, Georg Weizsäcker, Matthias Wibral, and especially Armin Falk for helpful discussions.

[*]Schaumburg-Lippe-Str. 5–9, 53113 Bonn, Germany. email:[last name at]iza.org

1 Introduction

The fungibility of money is a central principle in economics. It means that any unit of money is substitutable for another. In the analysis of consumer choice, for example, fungibility implies that consumption decisions are based on the consumer's total wealth—the specific composition of wealth is irrelevant (Modigliani & Brumberg 1954). Usually, economic theory implicitly assumes money to be fungible.

However, a number of studies has demonstrated that assuming a lack of fungibility can explain several empirical phenomena that are hard to reconcile with standard economic assumptions. Examples come from the fields of stock trading (Odean 1998, Barberis & Huang 2001), asset pricing (Benartzi & Thaler 1995, Barberis et al. 2001), stock market participation (Barberis et al. forthcoming), and life-cycle saving (Shefrin & Thaler 1988).[1] Yet, surprisingly little is known about the degree to which individual decision-making is in line with the notion of fungibility.

In this paper, we test experimentally whether consumers treat different forms of wealth as fungible.[2] In the experiment, subjects make a consumption decision over two goods. We induce a standard microeconomic utility function by specifying monetary payoffs for the possible consumption bundles. Subjects' budget consists of two components: an initial endowment in cash and a subsidy payment. We investigate two treatments, where the only treatment difference is the form of the subsidy. In the baseline treatment, called *Cash treatment* (CT), the subsidy is an unconditional cash payment. As both wealth components are cash and as the subsidy is not explicitly linked to either consumption good, consumers should realize that any unit of the subsidy payment is a perfect substitute for any unit of the initial endowment. In the main treatment, called *Label treatment* (LT), the subsidy is given as an in-kind benefit, i.e., subjects are required to spend the entire subsidy on the subsidized good. Still, the parameters are chosen such that the in-kind benefit is

[1] A related phenomenon is the so-called flypaper effect (e.g., Hines & Thaler 1995). However, the setting investigated in our paper is one of individual decision-making, whereas the flypaper effect is typically seen as the outcome of a strategic interaction between several agents.

[2] The following analogy from physics illustrates the basic idea behind our approach (Note that a violation of fungibility implies that consumers have some sort of cognitive or mental accounts (Thaler 1985) between which money can not be shifted): assume that you have two containers filled with water to the same level. You hypothesize that the two containers are linked to each other. A direct test of this hypothesis is to fill water into one of the containers. If they are linked, the water level in the two containers will adjust to equality again—as in the well-known communicating-vessels experiment. In contrast, if the two containers are not linked, the water level in the container to which the water was added will be higher than the water level in the other container.

non-distortionary: the amount of the subsidy is lower than the amount spent on the subsidized good in the (first-best) optimum. Through shifting the remaining budget, subjects can reach the same optimal consumption level as in the Cash treatment. Thus, the difference between treatments boils down to a mere difference in the label that is attached to the subsidy. If fungibility holds, subjects should treat any unit of the in-kind benefit as a perfect substitute for any unit of the initial endowment. Standard theory therefore predicts consumption to be the same across treatments. If we find, in contrast, that consumption of the subsidized good is higher in the Label treatment, we can conclude that consumption behavior is not in line with the principle of fungibility.

The experimental results show indeed that fungibility does not hold in the setting under investigation: whereas 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. The average marginal propensity to consume the subsidized good out of the subsidy is .280 in the CT vs .574 in the LT.[3] Moreover, decisions are heterogenous: 25% of subjects in the LT even have an MPC of 1.0. Most of the treatment difference is driven by these subjects, but the remaining subjects are also influenced to a certain extent by the treatment manipulation.

Remarkably, the heterogeneity of subjects' decisions is systematic: in the group of subjects with relatively strong mathematical skills (according to their high school math grade), behavior is consistent with standard economic theory. By contrast, in the group of subjects with weaker mathematical skills, the treatment difference is large and significant, which supports the view that the violation of fungibility occurs due to cognitive reasons. Moreover, we demonstrate that the treatment difference is *not* caused by feelings of moral obligation that might lead the recipient to comply with the label on the subsidy. These findings relate our study to recent work by Frederick (2005), Benjamin & Shapiro (2005) and Casari et al. (forthcoming), who show that people with higher 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.

In addition to the laboratory experiment, we conducted a vignette survey. The survey is methodologically complementary to the experiment. The lab experiment ensures incentive compatibility and allows us to collect extensive data about the subjects in a post-experimental questionnaire. The survey, in turn, enables us to check the robustness of our results in a less abstract environment. While the general

[3]The increase in consumption and the MPC are measured against consumption in a stage preceding the main stage described above. In this first stage, subjects only have the cash amount without the subsidy at their disposal.

structure of the survey is analogous to the structure of the experiment (we have again the same two treatments, CT and LT), the scenario (or the "vignette") is designed to be as close as possible to participants' real-life experience. Moreover, participants in the vignette survey choose only from two options, which makes the decision problem straightforward and is likely to reduce error rates.[4] The results of the survey confirm the experimental findings: participants in the Label treatment consume significantly more of the subsidized good and this effect is more pronounced among respondents with a relatively weak mathematical background.

Taken together, our results show that there is reason to assume that consumers do not always act in line with fungibility. Theories of life-cycle saving and stock market behavior should take this into account (see, e.g., Shefrin & Thaler 1988, Barberis & Huang 2001). Furthermore, the specific design of our study allows us to explain observed behavior of recipients of in-kind benefits. Under the assumption of fungibility, in-kind benefits do not distort the decision of inframarginal recipients, i.e., of recipients who would spend more than the subsidy on the subsidized good if they received a cash payment of the same amount (Moffitt 1989). If consumers do not always treat money as fungible, in-kind benefits will influence consumption more than previously thought, as they distort also the consumption decision of inframarginal consumers. Such a strong impact has been documented for housing benefits (Susin 2002, Gibbons & Manning 2006, Fack 2006). Taking our experimental results at face value, tenants who receive housing benefits will have a higher willingness to pay for a given apartment than tenants who get the same amount as cash grant. If landlords realize this behavior, they can seize an increase in housing benefits to increase the rent. We discuss this issue in more detail in Section 5.

A few papers investigate fungibility in laboratory experiments, but in environments that are quite different from ours. The studies by Gneezy & Potters (1997) and Thaler et al. (1997) address the interplay of fungibility and loss aversion for decisions about subsequent, identical gambles. In our setting, there is no uncertainty involved and losses are ruled out. Rockenbach (2004) examines fungibility in a binomial option-pricing task. She finds that investors do not realize that the value of a call option in such a setup is independent of the probability with which the good state of the world occurs. This behavior is not in line with fungibility, but with investors evaluating safe and uncertain investments separately.

[4]Compared to field data, experiment and survey both have the advantage to prevent self-selection into treatments and other confounding effects like intra-household bargaining. More specifically concerning in-kind benefits, we know for certain whether a recipient is inframarginal and we can make sure that there are no differences in the characteristics of the benefit payments. All these factors are discussed as reasons why field evidence on fungibility comes to mixed conclusions (see, e.g., Kooreman 2000 and Blow et al. 2004).

But why should people treat money as non-fungible? Tversky & Kahneman (1981) claim 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, Rabin & Weizsäcker 2006). 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 experimental finding that subjects with lower cognitive skills violate fungibility more often points also to narrow bracketing as a potential explanation for the treatment effect. The related notion of "mental accounting" (Thaler 1985, 1999) proposes that consumers use 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 (Heath & Soll 1996, Prelec & Loewenstein 1998).[5]

The remainder of the paper is organized as follows: The design of the experiment is described in Section 2. Section 3 presents experimental results and analyzes explanations for our findings. Section 4 reports design and results of the vignette survey and Section 5 contains an application of our results to benefit payments. Section 6 concludes.

2 Experimental Setup

2.1 Experimental Design

In the experiment, subjects have to make two subsequent consumption decisions. 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 both consumption decisions, subjects can allocate their budget on two goods. For each good, we induce a standard microeconomic utility function by specifying monetary payoffs for the possible consumption levels. Total payoff is the sum of the payoffs for each of the two goods in reference and subsidy stage.

In the reference stage, subjects get a cash budget of 50 money units which they can allocate freely on the two goods. In the subsidy stage, subjects have again the

[5]Note that mental accounting can be an advantageous heuristic, e.g., when it serves the decision maker to overcome his self-control problems. In the setup we analyze in this study, however, deciding according to mental accounting or narrow bracketing is a cognitive mistake that leads to a lower payoff.

endowment of 50 money units at their disposal and additionally receive a subsidy payment of 30 money units. The only difference between the two treatments is the form of the subsidy in this second stage—the reference stage is identical in both treatments. In the *Cash treatment* (CT), the subsidy is given as an unconditional cash grant. The CT serves as our baseline treatment. In the *Label treatment* (LT), the subsidy is given as an in-kind benefit, i.e., the money has to be spent entirely on the subsidized good. Still, the parameters are chosen such that the in-kind benefit is *non-distortionary*. By shifting the remainder of their budget appropriately, subjects can reach the same optimal consumption level as in the CT. The difference between treatments therefore concerns only the label attached to the subsidy.

The exact specification of the payoff function is presented in Table 1. For each good, payoff is increasing in consumption and marginal payoff is weakly decreasing. For clarity, we call the two goods 'subsidized good' and 'other good' throughout the paper, although strictly speaking the subsidized good is only subsidized in the second stage of the Label treatment. Prices per unit are $p_s = 3$ for the subsidized good and $p_o = 2$ for the other good. The payoff function and the prices stay constant over the course of the experiment. Budget left over in any stage can neither be saved nor does it yield payoff. There is no time limit for the decisions.

In order to make the difference between the initial endowment and the subsidy payment more salient, subjects earn their endowment in a real-effort task. Before the consumption decisions are taken, subjects have to count the number of zeros in large spreadsheets that consist of zeros and ones. When they manage to determine the correct number of zeros in a given amount of time they earn 100 money units that are later split in half for the two consumption decisions.[6] 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).

2.2 Behavioral Predictions

Our goal in this paper is to test experimentally whether individual behavior is in line with fungibility. To fix ideas, consider the indifference curve diagram in Figure 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

[6]The precise rules are as follows: subjects get 8 large tables with 300 entries each. To complete the task, they have to count the correct number of zeros on four sheets within 15 minutes. An answer is also counted as correct if the number reported differs only by 1 from the true number. If subjects do not complete the task, they only get an endowment of 10 money units.

Consumption	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
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

Consumption	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
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 1: *Payoff functions used in the laboratory experiment. "Subsidized good" denotes the good that is subsidized in the subsidy stage of the Label treatment. Payoff points were converted into real money after the experiment.*

to $p_s = 1$. Assume that the consumer has a cash budget of size R at his disposal and that payoff is maximized at consumption bundle A. This corresponds to the situation in the reference stage. Regardless of whether the consumer treats different income components as fungible, he should choose A since fungibility is not needed for this decision.

In the subsidy stage, however, the predictions of the standard model differ from the predictions of a model that allows for a lack of fungibility. Let us start with the standard model. In the Cash treatment, the consumer receives an additional subsidy of size S in cash and the budget constraint shifts to the right (dashed line). Now the optimal consumption bundle is B. In the Label treatment, the subsidy is paid in-kind and the consumer faces a kinked budget constraint (solid line). Still, the kink does not affect optimal decision making. As the amount of S is lower than the amount s^B spent on the subsidized good in the optimum, the first-best choice B is still feasible. Therefore, under the standard assumption that fungibility holds, consumption should 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 can not be shifted. A related concept that includes non-fungible sub-budgets as a building block is the concept of "mental accounting" (Thaler 1985). However, the literature on mental accounting has little to say on the exact process of allocating different income components to mental accounts. In this paper, 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

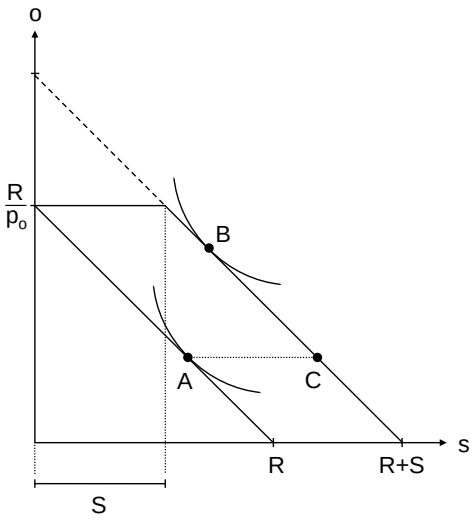


Figure 1: *Consumption decision with non-distortionary in-kind benefit. The subsidized good (s) is on the horizontal axis, the other good (o) is on the vertical axis. The dashed line is the budget constraint when the subsidy is given in cash. The solid line is the budget constraint when the subsidy is given in kind.*

occurs in the Label treatment. The consumer will allocate the cash endowment optimally (bundle A). However, the whole subsidy 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. This results in a consumption of bundle C where $s^C = s^A + S$ (see Figure 1). If both goods are normal, $s^C \geq s^B$.[7]

Therefore, if some subjects do not treat money as fungible, we should expect average consumption 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.

For the actual parameters in our experiment, the optimal consumption bundle in the reference stage is $A = (12, 7)$. The optimal bundle in the subsidy stage is $B = (13, 20)$, while bundle C is $(22, 7)$.

[7]Note that this reasoning depends on the order in which cash budget and subsidy are decided upon, especially, whether the decision in the reference stage (bundle A) is taken as given when deciding over the allocation of the subsidy payment. If the consumer spent the subsidy first, he would be able to allocate the cash budget so as to reach bundle B. In the experiment, we are therefore testing the joint hypothesis of fungibility and order of spending. However, the reference stage should lead most subjects to first allocate the cash endowment in the subsidy stage and then the subsidy, as they have already calculated how to allocate the cash budget.

2.3 Procedural Details

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 complete a number of example 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 1999). 92 subjects participated in the experiment, of whom one subject failed to complete the real-effort task. This leaves us with 45 independent observations in the Cash treatment and 46 observations in the Label treatment. Payoff points (cf. Table 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.

2.4 Vignette Survey

To test the robustness of our findings in a less abstract environment, we complement our laboratory analysis by conducting a vignette survey. The general structure is analogous to the structure of the experiment: participants decide twice, in a reference stage and a subsidy stage, and there are two treatments, a Cash treatment and a Label treatment.[9] The survey is designed such that participants can only choose between two options in each stage. This makes decisions much easier and should reduce error rates. Most importantly, we do not induce a utility function but rely on the true preferences of participants about the goods at hand. A total of 525 students participated in the survey. None of the respondents participated in our experiments. More details on the survey will be presented in Section 4.

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

[9] Since participants decide only hypothetically they do not have to complete a real-effort task before deciding.

3 Experimental Results

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 ability. Finally, we demonstrate that subjects' moral attitudes cannot explain the treatment effect.

3.1 Consumption in the Subsidy Stage

We focus on the results in the subsidy stage of the experiment, as this is the main treatment stage. Our first result concerns the impact of the labeled subsidy on consumption choice.

> **Result 1:** *Consumption of the subsidized good is significantly higher in the Label treatment. The marginal propensity to consume out of the subsidy is twice as large in the Label treatment as in the Cash treatment.*

A histogram of consumption choices in the two treatments is shown in Figure 2. 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 most frequent choice is to consume the optimal amount of the subsidized good (13 units), and average consumption is 14.4 units.[10] By contrast, the most frequent choice in the Label treatment is a consumption of 20 units, and only few 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 treatment difference is highly significant (Mann-Whitney U-test, $p = 0.003$, one-sided). Moreover, subjects in the Label treatment leave money on the table, as their choices translate into significantly lower payoffs (U-test, $p = 0.007$, one-sided).

To compute the marginal propensity to consume out of the subsidy payment, we compare decisions in the subsidy stage to decisions in the reference stage. A histogram of the intra-person change in consumption is shown in Figure 3. On average, the consumption change is +5.7 units in the Label treatment, compared to +2.8 units in the Cash treatment.[11] This difference is highly significant (U-test,

[10] Very few subjects in the subsidy stage choose a consumption bundle that is not on the Pareto frontier (1 in the LT and 2 in the CT). 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 when we confine the analysis to the Pareto optimal choices.

[11] Note that the design of the reference stage is exactly the same in both treatments. In partic-

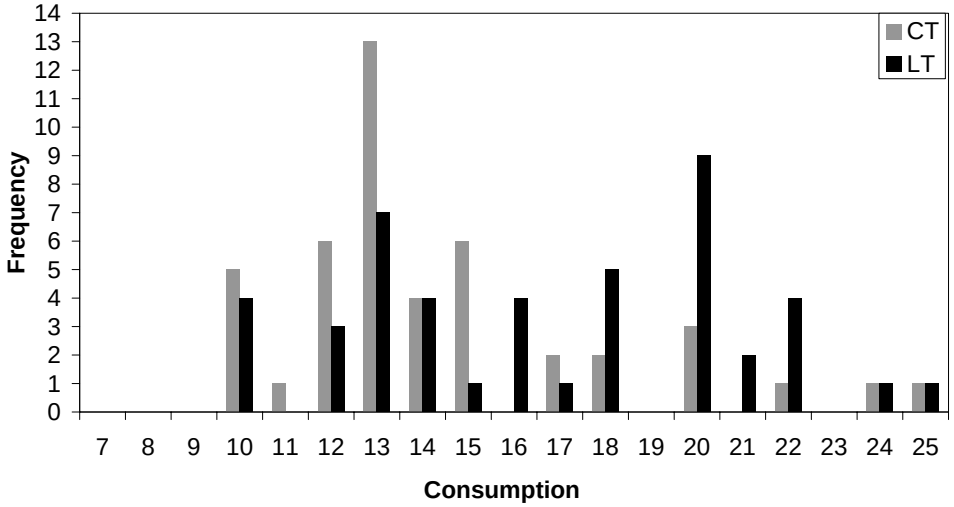


Figure 2: *Consumption of the subsidized good in the subsidy stage.*

$p < 0.001$, one sided). As the subsidy payment has a value of 10 units of the subsidized good and as the underlying cash budget remains the same, the resulting marginal propensity to consume out of the subsidy is .574 in the Label treatment vs. .280 in the Cash treatment.

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

> **Result 2:** *Most of the treatment difference is driven by subjects who increase their consumption by the full amount of the subsidy.*

As Figure 3 shows, the modal choice in the Label treatment is a consumption increase by 10 units. This decision implies that the entire subsidy is spent on the subsidized good, on top of the consumption from the reference stage. Subjects who treat the income sources as completely non-fungible will do exactly this (cf. bundle C in Figure 1). In the Label treatment, 10 out of 46 subjects spend the whole subsidy on the subsidized good, while this is true for only 1 out of 45 subjects in the Cash treatment. These subjects drive most of the treatment effect, but not all of it. If

ular, subjects are not aware of the fact that there will be two different treatments in the subsidy stage. We find that decisions are very similar across treatments: on average, subjects in the Label treatment buy 11.0 units of the subsidized good, while subjects in the Cash treatment buy 11.6 units. These results are very close to each other, but the difference is marginally significant (U-test: $p = .066$, two-sided). Still, we are not too worried about this marginal difference. First, in terms of profits, there is virtually no treatment difference (U-test: $p = .983$, two-sided). Second, consumption is slightly *lower* in the Label treatment. Thus, if there is any inertia in subjects' decisions, results from the reference stage work against the treatment effect we find in the subsidy stage, making our results stronger.

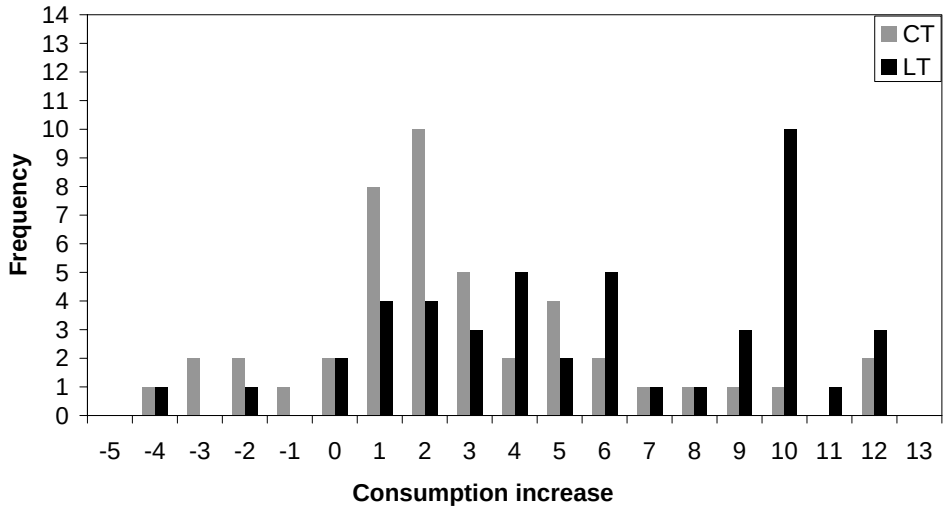


Figure 3: *Consumption increase of the subsidized good from reference stage to subsidy stage. The subsidy is worth 10 units of the subsidized good.*

we exclude these 11 subjects from the analysis, the treatment difference in absolute consumption remains. It becomes smaller (1.4 units, previously 2.3 units), but is still significant (U-test: $p = .044$, one-sided). The same is true for the treatment difference in consumption change (1.9 units, previously 2.9 units, $p = .001$).

Interestingly, subjects who spend the entire subsidy on the subsidized good seem to understand the general task in the experiment quite well: in the reference stage, they consume slightly less of the (later to be) subsidized good compared to all other subjects (10.3 vs 11.4 units), but the difference is not significant (U-test: $p = .141$, two-sided). In the subsidy stage, they 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: $p < .001$, one-sided). 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 being derived from 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: $p < .001$, one-sided) and also less than the other subjects in the Label treatment ($p = .002$).

3.2 Impact of Mathematical Ability

A consumer who does not treat different income components as fungible reduces the complexity of his consumption decision. In our setup, ignoring fungibility divides the rather complex two-good decision into two simple one-dimensional problems.

Subjects with lower cognitive and 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 1 that a consumer who brackets his decisions narrowly, i.e., who does not decide globally, will violate fungibility. Read et al. (1999) conjecture that narrow bracketing is negatively correlated with cognitive ability.[12] Our next result supports their intuition:

> **Result 3:** *The treatment difference in consumption is larger for subjects with lower mathematical ability.*

To analyze the interplay of mathematical ability and the treatment effect we divide the sample into a group with higher mathematical ability ("High-Math") and a group with lower ability ("Low-Math"). Since it could be that mathematical ability impacts not only the decision in the Label treatment but also in the Cash treatment, we compare the treatment difference within the High-Math group to the difference within the Low-Math group. Our proxy for mathematical ability is subjects' math grade in their high school completion exam. Table 2 shows the math grades of our subjects, where grades range from 1 ("excellent") to 5 ("fail"). The German high school system has one particularity: two years before the final exam, students have to choose two intensive courses replacing the respective standard courses.[13] 44% of subjects in our sample had a math intensive course in high school.

It is safe to assume that the average mathematical ability in a math intensive course is higher than in a standard course. At the same time, there is ample reason to suppose that there is some overlap in mathematical ability across the two courses, i.e., not every student in an intensive course has a higher mathematical ability than every student in a standard course.[14] To make grades in intensive and in standard

[12] 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 however no role. Subjects with lower cognitive ability should be less likely to realize this fact and should thus be more likely to use this heuristic.

[13] A typical intensive course encompasses 6 hours per week, whereas a typical basic course encompasses only 3 hours per week. Moreover, grades in the intensive course receive larger weights when computing the grade point average.

[14] There are several reasons for why sorting according to mathematical ability into the intensive course may not be perfect. For instance, students have to choose exactly two courses. If the correlation between different course-related abilities is very high, i.e., if a student is good (or bad) at everything, the choice of intensive course is no longer informative. Moreover, when students choose their intensive courses, they might misjudge their own mathematical ability. Finally, schools often have to ration the choices since it is not possible to offer all students the desired combination of intensive courses. For example, one of the authors (JA) wanted to choose History as intensive

Table 2: *Frequencies of math courses and grades*

Grade	Intensive course	Standard course
1	18	10
2	9	13
3	7	13
4	5	12
5	1	3

Notes: 1 is the best grade, 5 means failed.

courses comparable, we therefore assume that the mathematical ability of a student in the intensive course is equivalent to the ability of a student in the standard course who has a grade which is better by some amount I. For example, if $I = 1$, a subject with grade 2 ("good") in the intensive course is assumed to have the same ability as a subject with grade 1 ("excellent") in the standard course.

Table 3 presents the results for $I = 1$, but our findings are quite robust to the choice of this parameter.[15] As we do not want to impose an ad-hoc threshold regarding the dividing line between the High-Math and the Low-Math group, we analyze the treatment effect for a variety of thresholds. Every row in Table 3 corresponds to one threshold. In the strictest specification of High-Math (top row), only subjects with math intensive course and the best possible grade are classified as High-Math. As one goes down the table, more and more subjects are classified as High-Math and the ability of the marginal subject gets lower and lower.

We find, independent of the threshold we choose, the treatment effect in the Low-Math group to be always larger than in the High-Math group (see last column of Table 3). Moreover, the treatment effect in the High-Math group is very small for narrow classifications of who is High-Math (1.1 units) and rises steadily as the classification is broadened (up to 1.8 units). In the Low-Math group, the picture is even more striking: if many subjects are classified as Low-Math, the treatment difference is 2.6 units and rises to 4.5 units for a narrow definition of Low-Math. Since the overall treatment effect is 2.3 units, our results imply that for all thresholds the High-Math group has a treatment effect lower than average, while the Low-Math group has a treatment effect higher than average.

The result we just presented for average choices is also apparent in the distri-

course and ended up with Physics.

[15] In Appendix A we present results for $I = 2$ and $I = 3$.

Table 3: *Treatment effects in consumption according to mathematical ability*

	N. Obs.	Treatment Effect		
High-Math Def.	(H-M/L-M)	High-Math	Low-Math	TE_{Low} - TE_{High}
IC: 1, SC: -	18/73	1.05 (0.156)	2.55 (0.005)	1.50
IC: 12, SC: 1	37/54	1.15 (0.123)	2.96 (0.009)	1.81
IC: 123, SC: 12	57/34	1.56 (0.041)	3.47 (0.016)	1.91
IC: 1234, SC: 123	75/16	1.80 (0.012)	4.50 (0.030)	2.70

Notes: A standard course grade is assumed to be equivalent to an intensive course grade worse by 1. The first column lists all grades for intensive (IC) and standard course (SC) which are considered as High-Math for this row. The treatment effect is the difference of consumption in the subsidy stage between LT and CT; a positive value indicates that subjects in the LT consume more. The p-value of a one-sided U-test between treatments is shown in parentheses. The last column depicts the difference in treatment effects between the Low-Math and the High-Math group; a positive value indicates a larger treatment effect for the Low-Math group.

bution of choices. Figure 4 shows cumulative frequencies of consumption of the subsidized good. The thresholds are chosen as in Table 3. The top panels show very narrow definitions of High-Math and the definition is broadened more and more for the lower panels. The left panel always depicts the High-Math group, the right panel the Low-Math group. We find for all thresholds that the two consumption distributions in the Low-Math group are further apart than in the corresponding High-Math group, i.e., the treatment effect is always larger for the Low-Math group. The same effect can also be seen within a column: The more subjects are classified as High-Math (bottom panels) the larger the treatment difference. The same is true in the right column: when the Low-Math group includes fewer subjects the treatment effect is larger.

The pattern according to mathematical ability can also be seen in other aspects of subjects' behavior. Subjects with lower mathematical ability tend to increase their consumption more than subjects with higher mathematical ability. Moreover, we find that in the subsidy stage, High-Math subjects earn more than Low-Math subjects, whereas in the reference stage, consumption and profits are not correlated with mathematical ability. We have shown before that subjects who spend the whole subsidy on the subsidized good account for most of the treatment effect. The relation between mathematical ability and consumption in the subsidy stage holds also for these subjects. For all thresholds in Table 3, the probability of subjects who spend the entire subsidy on the subsidized good being Low-Math is higher than the probability of the other subjects.

3.3 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 alternative explanation is in line with evidence on the importance of reciprocity (see, e.g., Falk & Fischbacher 2006). In our context, the recipient would want to reciprocate the friendly action of the benefit giver by complying to the label attached to the subsidy. The next result shows that behavior in the experiment is not related to subjects'

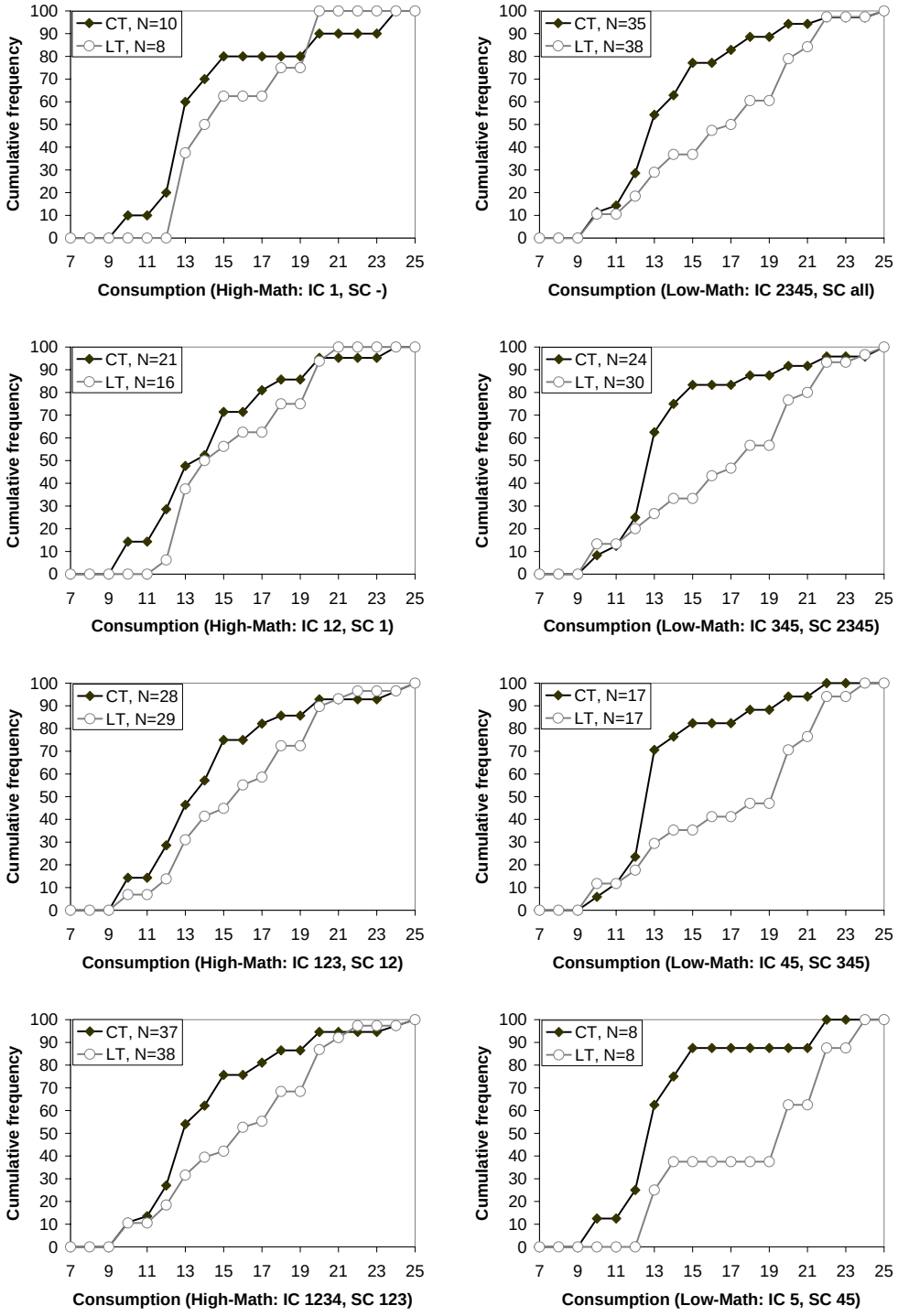


Figure 4: *Cumulative distribution of the consumption of the subsidized good in the subsidy stage for High-Math (left column) and corresponding Low-Math groups (right column). Thresholds between High-Math and Low-Math are chosen as in Table 3. The black line always depicts consumption in the Cash treatment, the grey line consumption in the Label treatment.*

attitudes towards moral obligation.

> **Result 4:** *A feeling of moral obligation to comply with the label on the subsidy cannot explain the treatment difference in consumption. If anything, subjects with weaker moral obligation consume more of the subsidized good.*

To measure subjects' attitude regarding moral obligation, we included three different scenarios in the post-experimental questionnaire in which subjects had to judge the behavior of a fictitious person. The scenarios involve a person claiming student support provided by the state although not being entitled to it, a person temporarily claiming unemployment benefits although having a new job already on the horizon, and parents spending child benefit payments not on child-related goods. The child-benefit scenario is closest to the decision in the experiment and reads as follows:[16]

> 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 have to indicate on a point scale from 1 to 6 how they judge the fictitious persons' behavior, 1 indicating "not justified at all" and 6 indicating "fully justified". 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.

Analogous to our analysis of mathematical ability, we split the sample according to the answer to this vignette in a group with strong moral obligation ("High-MO") and a group with weak moral obligation ("Low-MO"). Again, we avoid concentrating on an ad-hoc threshold by analyzing different specifications for the High-MO and the Low-MO groups. The results are depicted in Table 4, with each row corresponding to one threshold. We find that the treatment effect is close to the overall treatment effect of 2.3 in most High-MO and Low-MO groups. In particular, there is no systematic increase or decrease in the treatment effect according to the strength of moral obligation. The last column indicates that for most thresholds the treatment effect is even stronger for subjects with *weaker* moral obligation. These results are not in line with the hypothesis that moral obligation drives the treatment effect in

[16]The two other scenarios can be found in Appendix B.

Table 4: *Treatment effects in consumption according to moral obligation*

	N. Obs.	Treatment Effect		
High-MO Def.	(L-MO/H-MO)	Low-MO	High-MO	TE_{High} - TE_{Low}
1	86/5	2.51	-0.50	-3.01
		(0.001)	(0.500)	
12	66/25	2.45	1.60	-0.84
		(0.008)	(0.096)	
123	53/38	2.09	2.33	0.24
		(0.028)	(0.059)	
1234	36/55	2.56	2.11	-0.45
		(0.031)	(0.042)	
12345	21/70	4.06	1.77	-2.29
		(0.010)	(0.050)	

Notes: Answers to the child-benefit scenario are the proxy for moral obligation. Low numbers correspond to a high moral obligation. The first column lists all answers which are considered as High-MO for this row. The last column depicts the difference in treatment effects between the High-MO and the Low-MO group; a positive value indicates a larger treatment effect for the High-MO group. For further details see Table 3.

the experiment. The answers to the two other vignettes are also not systematically related to the consumption of the subsidized good (see Appendix B). The same holds true if we take the average answer of each subject to the three vignettes as our proxy for moral obligation.

We have shown in Section 3.1 that most of the treatment difference is driven by subjects who spend the whole subsidy on the subsidized good in addition to their consumption in the reference stage. These subjects, however, do not differ in their moral obligation from other subjects. Their moral obligation is even a little bit lower but this difference is not significant for any of the three scenarios or the average answer to all three scenarios.[17]

[17]For example, the average answer in the child benefit scenario is 4.4 for subjects who spend the full subsidy on the subsidized good and 3.8 for the other subjects, indicating that the other subjects have a somewhat stronger moral obligation (low numbers indicate strong moral obligation). This difference is not significant (U-test: $p = .298$, two-sided).

4 Vignette Survey

In this section we report the results of the vignette survey. The survey is complementary to the experiment as it embeds the consumption decision in a more natural environment: we do not induce a utility function but rely on participants' true preferences. The scenario consists of an apartment choice task which is very similar to the housing decisions a typical participant faces in life.[18] Compared to the experiment, the choice set is much smaller: participants choose only between two options. The survey is thus a robustness check whether results of the experiment generalize to a simpler and less abstract environment.

4.1 Design

The survey has the same general structure as the experiment. Participants decide twice, in a reference stage and in a subsidy stage, and there are two treatments, the Cash treatment (CT) and the Label treatment (LT). In each stage, the survey presents participants with a scenario in which they have to decide hypothetically which of two apartments to rent, either a small and cheap one (250 Euro per month) or a large and expensive one (410 Euro per month).

In the reference stage, participants' only income source is an unconditional cash grant of 900 Euro per month. In the subsidy stage, participants are offered the same apartment choice again, now with a budget of 1100 Euro per month. The participants have to indicate which apartment they would choose in the two situations. See Appendix D for the precise wording of the questionnaire. While the reference stage is exactly the same in both treatments, the subsidy stage differs in the way the additional payment is labeled: in the CT, participants' initial budget is increased by a cash scholarship of 200 Euro, whereas in the LT, the budget is increased by a housing allowance of 200 Euro that has to be spent on rent expenditures. Yet, as the amount of the payment is lower than the rent of the cheaper apartment, the housing allowance is non-distortionary.[19]

The treatments were assigned randomly and nobody participated in more than one treatment. Participants were students of the universities of Bonn and Cologne. Overall, we collected the choices of 525 participants, 264 in the Cash treatment and 261 in the Label treatment. None of the participants took part in the laboratory

[18] The figures used in the scenario are close to actual average budgets and rent expenditures of German students, according to a representative survey of the German student population on behalf of the Federal Ministry of Education and Research (Isserstedt et al. 2004).

[19] Some participants in the LT even wrote a remark on the survey sheet, noting that the allowance is not binding.

experiment. In addition to the apartment choices, we asked participants to indicate their age, sex, and field of study.

If participants treat the two income sources as fungible, the share of participants choosing the large apartment should be identical across treatments; but if some participants violate fungibility, we expect average housing consumption to be higher in the Label treatment.

4.2 Consumption in the Subsidy Stage

The main research question for the survey is whether the label on the additional payment influences reported consumption decisions. Choices in the subsidy stage show that this is indeed the case:

> **Result 5:** *Participants in the Label treatment choose the large apartment significantly more often than participants in the Cash treatment.*

We first consider the full sample. In the Cash treatment, 51.5 percent of participants choose the large apartment, whereas in the Label treatment 62.8 percent of participants do so. The difference is highly significant (Fisher exact test, $p = 0.006$, one-sided). This comparison, however, understates the true effect since the full sample includes also participants who choose the large apartment already in the reference stage (roughly 22 percent of the sample).[20] For these participants we cannot observe whether the treatment manipulation has an effect on their choice, as they cannot increase their housing consumption further when having the larger budget in the subsidy stage. To get a more precise picture of the impact of the subsidy, we exclude all participants from the sample who chose the large apartment in the reference stage, leaving us with 408 observations. In this smaller sample, 39.2 percent of participants in the CT switch to the large apartment in the subsidy stage. The label on the subsidy increases the probability to switch by 13.5 percentage points: 52.7 percent of participants in the LT switch to the large apartment. This marginal effect of 34 percent is highly significant (Fisher exact test, $p = 0.004$, one-sided). The marginal propensity to consume out of the subsidy jumps from 0.314 in the CT to 0.422 in the LT.[21]

[20] In the reference stage of the CT, 20.8 percent of the participants choose the large apartment, whereas this is true for 23.8 percent in the LT. This difference is not significant (Fisher exact test, $p = 0.242$, two-sided).

[21] Switching from the small to the large apartment increases housing consumption by 160 Euro (410 Euro - 250 Euro), and the subsidy is worth 200 Euro.

4.3 Impact of Mathematical Ability

In the experiment, subjects with lower mathematical ability were influenced more strongly by the treatment manipulation. The same holds true for the survey:

> **Result 6:** *The treatment effect is stronger for participants with a weaker mathematical background*

To test this hypothesis with our data, we use the major field of study as a proxy for participants' arithmetic skills. To rank the majors accordingly, we use data on the performance of test takers in the quantitative section of the Graduate Record Examination (GRE); in this section, test takers have to answer a large number of mathematical questions in a short amount of time.[22] We do not have test scores for the participants in our survey, but the average test scores for all test takers between 2001 and 2004. Therefore, the scores can serve only as a rough measure for the degree of mathematical reasoning in the respective field of study. Still, we are confident that the resulting ranking gives a meaningful picture of the degree of mathematical reasoning with which the respective majors are taught at German universities. Table 5 lists majors present among our participants ranked according to their average (quantitative) GRE score.

We split the sample according to GRE score in a High-Math and a Low-Math group. In order to not impose an ad-hoc threshold, we report results for four thresholds (550, 600, 650 and 700 points).[23] Table 6 reports results, each row corresponds to one threshold. The upper panel shows results for the full sample, the lower panel for the sample excluding participants who chose the large apartment already in the reference stage. The last column shows the difference in treatment effects between the Low-Math and High-Math group; a positive value of x indicates that the treatment effect in the Low-Math group is x percentage points larger than in the High-Math group. For the full sample, the treatment effect is indeed larger in the Low-Math group for three out of the four thresholds; for the restricted sample, this holds true for all thresholds. The data is noisier than the corresponding figures for the experiment (see Table 3) but this is what one would expect given the rough proxy for mathematical ability we use in the survey.[24]

[22] Prospective graduate students have to take the GRE test before applying to Ph.D. programs in the U.S., the UK, or Canada. For details concerning the GRE test, see www.gre.org. Data was provided by ETS, the institution organizing the test.

[23] The largest gap in test scores between two adjacent fields with at least 5 observations is between Chemistry (696 points) and Philosophy (636 points). This points to the threshold at 650 points.

[24] We can say little about whether moral obligation drives the results in the survey as we do not have a proxy for participants' feelings of moral obligation. However, the choice patterns along participants' mathematical background speak against a strong impact of moral obligation. It would

Table 5: *GRE test scores*

Field of Study	Average GRE Score	N. Obs.
Physics	744	35
Mathematical Sciences	733	6
Economics	706	26
Computer Sciences	704	51
Chemistry	696	46
Philosophy	636	5
Geology	630	11
Biology	615	51
Business	595	55
Geography	578	12
Foreign Languages	573	37
History	556	10
Medical Sciences	552	6
Sociology	545	11
Psychology	543	9
Law	539	85
Education	534	10
Communications	533	7

Notes: "Average GRE Score" refers to the mean score in the quantitative section of the GRE that test takers in the years 2001–2004 achieved, according to field of study. "N. Obs." refers to the number of participants in our sample who are enrolled in the respective field of study. Only fields of study with at least 5 observations are included in this table.

Table 6: *Apartment choices according to mathematical ability*

	N.Obs.	Treatment Effect		
Threshold	H-M/L-M	High-Math	Low-Math	TE_{Low} - TE_{High}
Full sample				
700	119/379	11.2	13.1	1.9
		(0.150)	(0.006)	
650	167/331	-2.2	20.1	22.3
		(0.450)	(0.000)	
600	241/257	11.9	13.3	1.4
		(0.042)	(0.022)	
550	368/130	13.0	11.4	-1.6
		(0.008)	(0.124)	
Participants choosing the small apartment in the reference stage				
700	91/294	11.4	16.0	4.6
		(0.189)	(0.004)	
650	121/264	4.3	21.2	16.9
		(0.390)	(0.000)	
600	180/205	14.2	16.0	1.8
		(0.038)	(0.016)	
550	289/96	13.2	18.8	5.6
		(0.016)	(0.051)	

Notes: The first column lists the threshold GRE scores between High-Math and Low-Math group. All participants with field of study having an average GRE score higher than the threshold are classified as High-Math. The treatment effect is the difference between LT and CT of the percentage of participants choosing the large apartment in the subsidy stage; a positive value indicates that participants in the LT choose the large apartment more often. The p-value of a one-sided Fisher exact test between treatments is shown in parentheses. The last column depicts the difference in treatment effects between the Low-Math and the High-Math group; a positive value of x indicates that the treatment effect in the Low-Math group is x percentage points larger than in the High-Math group.

5 Application to Benefit Payments

The specific design of our experiment additionally allows us to explain observed behavior of recipients of in-kind benefits. Contrary to the standard economic view, our findings suggest that in-kind benefits can lead to distortions even among infra-marginal consumers, as long as these do not act in line with fungibility. A straightforward application of this finding concerns the effect of housing benefits on market rents. In many countries, low-income tenants are subsidized through housing benefit payments. Empirical studies for the U.S. and for France have shown that the introduction of more generous housing benefits 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.

In contrast, 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 than tenants who get 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, they show that 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 constitute a transfer payment from taxpayers to landlords. To prevent landlords from exploiting that tenants treat money as non-fungible, one should try to reduce this behavior of tenants. We believe that the effectiveness of housing benefit payments can be improved by making a few simple policy changes. For instance, it is important to link the housing benefits less saliently to the rent payments by changing the periodicity of the benefit payments to differ from the periodicity of the rent payments. Similarly, housing benefits should be paid directly to the tenant and not to the landlord. In the French housing benefit system, for example, the subsidy is often paid directly to the landlord.[26]

---

be difficult to argue that these patterns are in fact due to systematic differences in moral attitudes.

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

[26] This might additionally affect the bargaining between landlord and tenant by creating a "moral property right" of the landlord (see Gächter & Riedl 2005). One could also let the exact amount of the subsidy depend on variables that the landlord cannot observe.

However, there are other benefit payments where it is good news if recipients violate fungibility. Child benefits, for example, are usually intended to increase spending on child-related goods. All recipients of child benefits are inframarginal since recipients are not restricted in their use of these funds and only the name of the benefit payment marks it as a separate income component. Still, if recipients do not treat money as fungible, they will spend more of the subsidy on child-related goods compared to if they received a cash grant of equal amount. Additionally, the supply side of child-related goods has a lower market power than landlords in the housing market. It is thus unlikely that suppliers of these goods can raise prices after increases in child benefits or discriminate between eligible and non-eligible consumers. Apparently, the effects we identify in the lab are also present in the field: Kooreman (2000) analyzes data from the Netherlands and demonstrates that the MPC for child clothing out of child benefits is much higher than out of the other income.[27] Munro (2005) investigates consumption behavior of households who receive a "winter fuel allowance" in the UK, an unconditional payment much like child benefits, and finds a positive effect on heating expenditures.

6 Conclusion

We pursue a dual research strategy by conducting an incentivized laboratory experiment and a randomized vignette survey to test whether consumers treat different income sources as fungible. Experiment and survey yield the same results: even in our simple setup, many subjects do not treat money as fungible. 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.

Our findings have important implications for several topics in economic research. Besides generating insights on the effects of benefit payments on consumption decisions, they are also related to research on stock market behavior and research on life-cycle saving. First, inframarginal consumers, i.e., consumers who receive an in-kind benefit but wanted to spend more than the subsidy anyway, might not be so inframarginal after all. They tend to comply with the label attached to the subsidy. Public policy can therefore influence consumers in a simple way by explicitly stating the intended use of the subsidy. We only address inframarginal consumers in this study, but distorted consumers might show a similar effect. If distorted consumers treat money as non-fungible, they will spend more than the amount of the subsidy

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

on the subsidized good, i.e., they will not consume in the kink of their budget constraint. Second, our findings lend support to field studies that explain behavior of stock market investors by assuming that investors are loss-averse and do not treat different wealth components as fungible (e.g., Benartzi & Thaler 1995, Odean 1998). These studies assume that investors evaluate their portfolio over some time interval and that they look at different components (e.g., stock vs. bond holdings) separately, i.e., that the different wealth components are separated over time and in the cross section. Whether consumers evaluate their investment over certain time periods was examined by Gneezy & Potters (1997) and Thaler et al. (1997); we provide a direct test of the second assumption by concentrating on cross-section fungibility. Third, our results are important for theories of life-cycle saving which usually rely on the assumption of fungibility. Since we find even in a very simple setup that 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). Finally, there is an additional application of our results in the field of savings. A large literature argues that people save too little for retirement (Mitchell & Moore 1998). One of the reasons cited is non-exponential discounting (Laibson 1997). Benjamin & Shapiro (2005) find that non-exponential discounting is negatively correlated with cognitive skills. We show that cognitive skills are also negatively correlated with a violation of fungibility. It might therefore be that people at the lower end of the skill distribution undersave due to non-exponential discounting and at the same time could be targeted by government interventions relying on people not treating income components as fungible. This hypothesis is however highly speculative; we will analyze it in future research.

We are confident that the effect we identified in the laboratory is indicative of behavior in everyday consumption decisions. First of all, the differences between income sources in our experiment were not very large. Although subjects had to fulfill a real-effort task to earn one part of their budget, both income parts were given to them by the experimenter. If people treat income sources as non-fungible already in our setup, we should expect this behavior to be stronger when differences between income sources are more pronounced, e.g., when the different wealth components are current wealth vs. future income or home equity vs. cash holdings. Additionally, in both experiment and survey, our subject pool consisted exclusively of university students. It is safe to assume that our subjects have higher cognitive skills than the population average. Since results were stronger for subjects with lower cognitive ability, we probably underestimate the true effect. Moreover, people at the bottom end of the skill distribution are also those who are more likely to receive public benefits since income is usually correlated with ability. They will thus be exposed more frequently to the specific institution we implemented in the laboratory.

Our results do not imply that everybody in every situation will violate fungibility. In our study, treating money as non-fungible is linked to cognitive ability 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 violation of fungibility is likely to become less frequent when subjects gain experience. Moffitt (1989) studies the food stamps program in Puerto Rico and finds no difference between (non-distortionary) in-kind benefits and cash grants. People in his study had considerable experience with food stamps and were apparently able to learn the rational behavior. By contrast, decisions on housing consumption are much less frequent; here, the scope for non-rational behavior is larger. In an extension to our experimental setup, one could analyze the interplay of learning and fungibility by repeating decisions or by explaining the concept of fungibility to subjects prior to the decision.

7 References

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 pricing', *Quarterly Journal of Economics* **116**(1), 1–53.

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

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. & Shapiro, J. M. (2005), 'Who is 'behavioral'? Cognitive ability and anomalous preferences', *Working Paper, Harvard University* .

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

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

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

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.

Fischbacher, U. (1999), 'z-Tree - Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual', *IEW Working Paper 21* .

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

Gächter, S. & Riedl, A. (2005), 'Moral property rights in bargaining with infeasible claims', *Management Science* **51**(2), 249–263.

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

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

Hines, James R, J. & Thaler, R. H. (1995), 'The flypaper effect', *Journal of Economic Perspectives* **9**(4), 217–26.

Isserstedt, W., Middendorff, E., Weber, S., Schnitzer, K. & Wolter, A. (2004), 'Die wirtschaftliche und soziale Lage der Studierenden in der Bundesrepublik Deutschland 2003', *Report for the Federal Ministry of Education and Research.* http://www.bmbf.de/pub/040714_ soz17hauptbericht.pdf.

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

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.

Mitchell, O. S. & Moore, J. F. (1998), 'Can Americans afford to retire? New evidence on retirement saving adequacy', *The Journal of Risk and Insurance* **65**(3), 371–400.

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* .

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.

Rabin, M. & Weizsäcker, G. (2006), 'Narrow bracketing and dominated choices', *mimeo* .

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

Rockenbach, B. (2004), 'The behavioral relevance of mental accounting for the pricing of financial options', *Journal of Economic Behavior & Organization* **53**, 513–527.

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

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

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., 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.

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

A Impact of Mathematical Ability

Table 7: *Consumption according to mathematical ability* ($I = 2$)

	N. Obs.	Treatment Effect		
High-Math Def.	(H-M/L-M)	High-Math	Low-Math	TE_{Low} - TE_{High}
IC: 1, SC: -	18/73	1.05 (0.156)	2.55 (0.005)	1.50
IC: 12, SC: -	27/64	0.76 (0.222)	2.87 (0.004)	2.11
IC: 123, SC: 1	44/47	1.96 (0.021)	2.57 (0.048)	0.61
IC: 1234, SC: 12	62/29	1.77 (0.024)	3.40 (0.029)	1.63
IC: all, SC: 123	76/15	1.90 (0.012)	4.29 (0.043)	2.39

Notes: A standard course grade is assumed to be equivalent to an intensive course grade worse by 2. For further details see Table 3.

Table 8: *Consumption according to mathematical ability* ($I = 3$)

	N. Obs.	Treatment Effect		
High-Math Def.	(H-M/L-M)	High-Math	Low-Math	TE_{Low} - TE_{High}
IC: 1, SC: -	18/73	1.05 (0.156)	2.55 (0.005)	1.50
IC: 12, SC: -	27/64	0.76 (0.222)	2.87 (0.004)	2.11
IC: 123, SC: -	34/57	1.92 (0.033)	2.52 (0.023)	0.60
IC: 1234, SC: 1	49/42	2.17 (0.012)	2.38 (0.085)	0.21
IC: all, SC: 12	63/28	1.88 (0.018)	3.19 (0.044)	1.31
IC: all, SC: 123	76/15	1.90 (0.012)	4.29 (0.043)	2.39

Notes: A standard course grade is assumed to be equivalent to an intensive course grade worse by 3. For further details see Table 3.

B Moral Obligation Vignettes

B.1 Student-Support Scenario

> Mr Smith is a first-year Biology student who wants to apply for Bafög.[28] 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? 1: "Not justified at all", 6: "Fully justified"

Table 9: *Consumption according to moral obligation (student-support scenario)*

	N. Obs.	Treatment Effect		
High-MO Def.	(L-MO/H-MO)	Low-MO	High-MO	TE_{High} - TE_{Low}
1	57/34	2.79	1.58	-1.21
		(0.004)	(0.103)	
12	34/57	1.59	2.50	0.91
		(0.119)	(0.008)	
123	22/69	2.19	2.10	-0.09
		(0.106)	(0.019)	
1234	8/83	1.25	2.37	1.12
		(0.385)	(0.004)	

Notes: For details see Table 4.

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

B.2 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? 1: "Not justified at all", 6: "Fully justified"

Table 10: *Consumption according to moral obligation (UI scenario)*

	N. Obs.	Treatment Effect		
High-MO Def.	(L-MO/H-MO)	Low-MO	High-MO	TE_{High} - TE_{Low}
1	62/29	1.88	3.23	1.35
		(0.024)	(0.030)	
12	50/41	2.75	1.68	-1.07
		(0.006)	(0.099)	
123	38/53	2.21	2.31	0.10
		(0.022)	(0.027)	
1234	30/61	1.93	2.44	0.51
		(0.069)	(0.010)	
12345	19/72	2.76	2.17	-0.59
		(0.059)	(0.010)	

Notes: For details see Table 4.

C Instructions of Laboratory Experiment

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

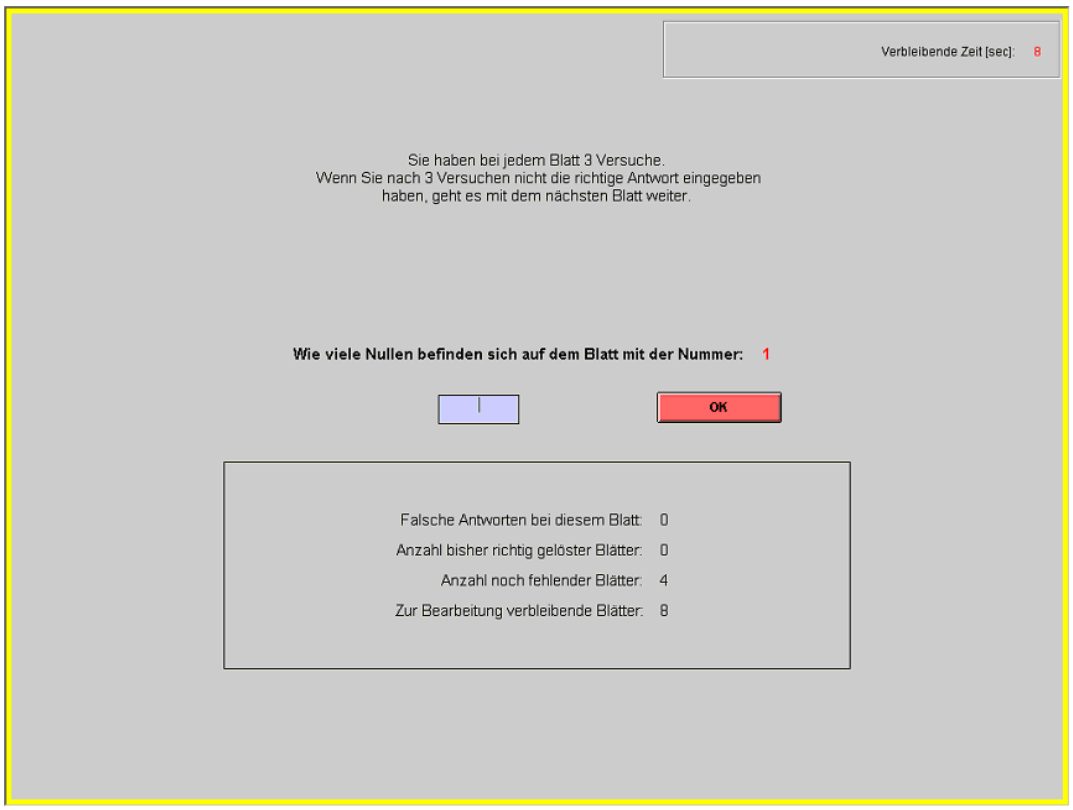
During the **work phase** you have the opportunity to earn 100 talers. The work consists of counting the number of zeros in tables filled with zeros and ones. Below, you see an example table with 3 rows and 8 columns. The tables used in the experiment are larger, they contain 10 rows and 30 columns.

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:

Total price of purchase =	(units of housing × price per unit of housing)
	+ (units of clothing × price per unit of clothing)

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

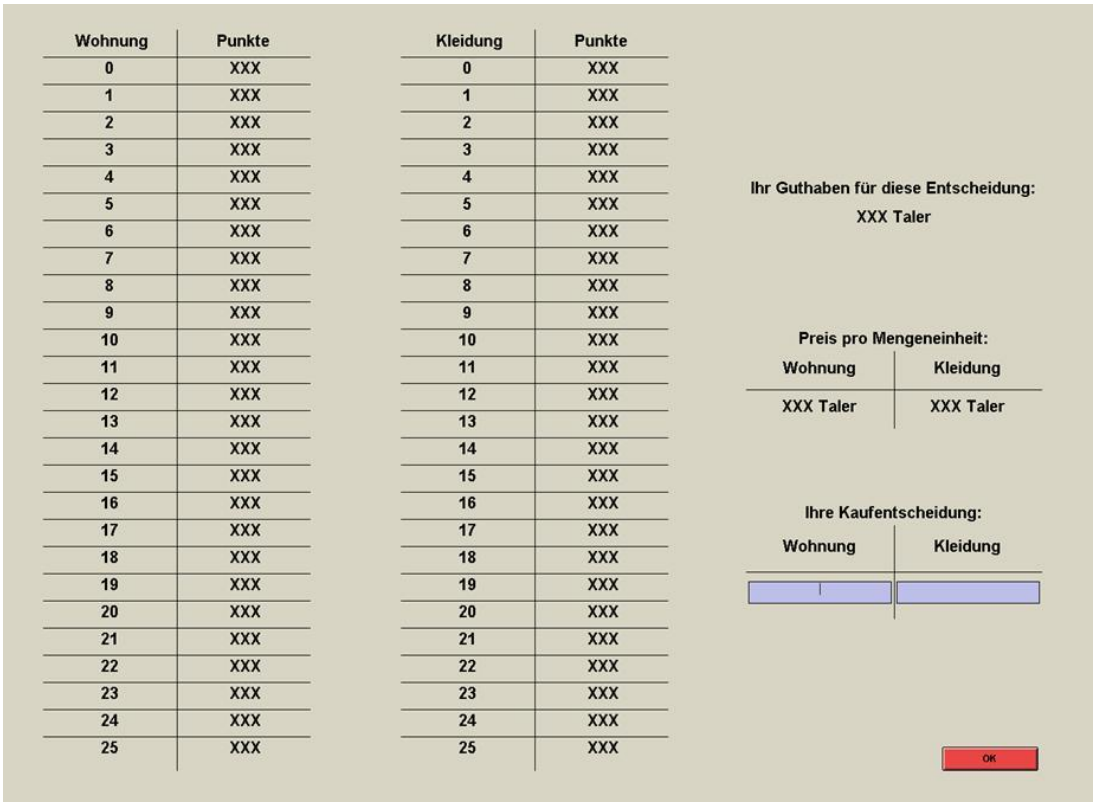
Example of a purchase

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

Label treatment

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

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

The **housing allowance** 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 allowance**.

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**.

D Exact Wording of the Survey

Hello!

We are conducting a survey as part of our doctoral research. Please imagine to be in the following situation and answer the two questions.

Suppose you have finished your studies and are about to start postgraduate studies in a new city. The university provides you with a monthly scholarship of 900 Euro. This is your only income. You have decided to live on your own. You find two apartments between which you can choose. Both are in the same building and are of similar standard. Apartment A has 25 sqm and the rent is 250 Euro per month. Apartment B has 37 sqm and the rent is 410 Euro per month. Which apartment would you rather choose in the financial situation described above?

☐ Apartment A ☐ Apartment B

[Cash treatment]: Now suppose that the city administration provides you with a monthly scholarship of 200 Euro, in addition to your university scholarship of 900 Euro. Which apartment would you rather choose in this new situation?

[Label treatment]: Now suppose that the city administration provides you with a monthly housing allowance of 200 Euro, in addition to your university scholarship of 900 Euro. This housing allowance has to be spent exclusively on rent expenditures. Which apartment would you rather choose in this new situation?

☐ Apartment A ☐ Apartment B

Field of Study: .....

Age: .....

☐ Female ☐ Male

Thank you!