Fungibility, Labels, and Consumption[†]

Johannes Abeler* **Felix Marklein**

University of Nottingham *University of Bonn*

July 7, 2010

Fungibility of money is a central assumption in the theory of consumer choice: any unit of money is substitutable for another. This implies that the composition of income or wealth is irrelevant for consumption. We find in a field experiment that even in a simple, incentivized setup many subjects do not treat money as fungible. When a label is attached to a part of their budget, subjects change consumption according to the label. A controlled laboratory experiment confirms this result and further shows that subjects with lower cognitive abilities are more likely to violate fungibility. The findings lend support to behavioral models of narrow bracketing and mental accounting. One implication of our results is that in-kind benefits distort consumption more strongly than usually assumed.

JEL classification: C91, C93, D01, H31, I38.

Keywords: Fungibility, In-kind Benefits, Mental Accounting, Narrow Bracketing, Field Experiment, Laboratory Experiment

---

[†]Financial support from the DFG (through GRK 629), the Bonn Graduate School of Economics, and the Nottingham School of Economics is gratefully acknowledged. We thank Sumit Agarwal, Stefano DellaVigna, Gabrielle Fack, Simon Gächter, Uri Gneezy, Lorenz Goette, Paul Heidhues, Steffen Huck, David Huffman, David Jaeger, Peter Kooreman, Anne Laferrère, David Laibson, Ulrike Malmendier, Wolfram Merzyn, Jörg Oechssler, Susanne Ohlendorf, Robert Oxoby, Hans-Theo Normann, Matthew Rabin, Marieke Schnabel, Lael Schooler, Bob Sugden, Uwe Sunde, Ian Walker, Georg Weizsäcker, Matthias Wibral, Ro'i Zultan, and especially Steffen Altmann and Armin Falk for helpful discussions and Susanne Nett, Daniel Nett, and Petra Gabler for helping to conduct the field experiment. Valuable comments were also received from numerous seminar and conference participants.

*Corresponding author: johannes.abeler[at]nottingham.ac.uk; University of Nottingham, School of Economics, Nottingham, NG7 2RD, UK

1 Introduction

A central assumption in the theory of consumer choice is the fungibility of money—any unit of money is substitutable for another. As a consequence, the composition of wealth (or income, respectively) is irrelevant for choices and consumption decisions are based on total wealth alone.

In contrast, several behavioral theories of decision-making argue that individual choices often violate fungibility. Theories of narrow bracketing (Tversky & Kahneman 1981, Barberis et al. 2006, Rabin & Weizsäcker 2009) are based on the assumption that people break down complex decision problems into several parts and decide on each part separately. Narrow bracketing predicts that consumers ignore background wealth or other income sources when deciding on how to spend, e.g., their labor income. Similarly, models of mental accounting (Thaler 1980, 1985, 1999) assume that consumers form mental budgets to organize their financial decisions. As money is not fungible across these mental budgets, choices can be constrained. Consumers might violate fungibility also because of feelings of reciprocity towards the provider of the income. The person or institution who provided the money might have clearly stated preferences about the final allocation of consumption. For example, many governments state explicitly that they care about the welfare of children when handing out child benefits. To reciprocate the kind act of the benefactor, the consumer might want to honor these stated preferences, change consumption accordingly, and thereby violate fungibility (for related models of reciprocity, see Rabin 1993, Falk & Fischbacher 2006).

An ideal way to investigate whether consumers treat money as fungible is to analyze their spending behavior when they receive targeted payments. In particular, consider the case in which a consumer receives a *non-distortionary* in-kind grant, i.e., an in-kind grant with an amount lower than what the consumer would have spent on the targeted good anyway. For example, one could think of a tenant who wants to spend \$500 on rent and receives housing benefits of \$200. Rational consumers, who treat money as fungible, will spend such an in-kind grant in exactly the same way as an unconditional cash grant: By shifting their remaining budget, they can comply with the condition of the grant and still reach the same first-best consumption level. The only difference to a cash grant is the label attached to the grant. By contrast, a consumer who does not treat the targeted payment as fungible will not use the possibility to reallocate parts of his original budget and will increase his consumption of the targeted good beyond the level of the first-best consumption bundle.[1]

Field evidence on how consumers spend in-kind benefits might therefore help to answer whether consumers treat money as fungible. Unfortunately, this field data is influenced by many factors that make it hard to clearly identify whether consumer behavior is caused by violations of fungibility or not. For example, increases in housing benefits for low-income tenants have lead to pronounced rent increases (e.g., Susin 2002, Gibbons & Manning 2006, Fack 2006). This effect could well be driven by a violation of fungibility, as tenants' willingness to pay for a given apartment is increased and landlords take advantage of this. However, these price increases might also be due to the low elasticity of housing supply. More direct evidence on fungibility comes from labeled cash grants like child benefits. Kooreman (2000) finds that the marginal propensity to consume child clothing out of child benefits is higher than out of other income, violating fungibility. But even for this kind of benefit it is debated how other factors such as intra-household bargaining or the characteristics of the benefit payment (e.g., periodicity) influence results (see Blow et al. 2007, Edmonds 2002).

In this paper, we use controlled experiments as an alternative and complementary way to investigate whether consumers treat money as fungible. Our experimental setup allows for clear identification of causality, since the type of benefit received is exogenously imposed by the experimenter. Moreover, we can investigate behavior in a tightly controlled environment where the factors confounding field data are excluded by design.

We pursue a dual research strategy by combining a field experiment and a laboratory experiment, both based on the same general design. Subjects decide over the consumption of two goods. In addition to their cash endowment, subjects either receive a non-distortionary in-kind grant or a cash grant of equal size. Both grants are paid lump-sum. Rational subjects who treat money as fungible should not be influenced by whether the grant is given as cash or in kind. In contrast, subjects who do not treat money as fungible will spend this grant disproportionately on the targeted good and thus consume too much of this good.

We chose a restaurant as setting for our field experiment because in this environment guests consume two distinct goods: they eat and they drink at least a minimal amount. Thus, an in-kind grant below this minimal amount does not distort the consumption decision. Guests received either a voucher for beverage consumption

[1] A similar phenomenon has been called a "flypaper effect" as the money "sticks where it hits" (e.g., Hines & Thaler 1995).

or a voucher for the total bill. They did not know that they participated in an experiment. Participants thus acted in a naturally occurring, incentivized, well-known environment and felt unobserved. They could not self-select into treatments or the experiment at all since vouchers came as a surprise to participants after they had entered the restaurant and since treatments were assigned exogenously. We find that an 8-euro beverage voucher increases beverage consumption on average by 3.90 euros compared to a bill voucher. The difference between treatments is thus very large, almost half of the treatment manipulation. We show that this effect is not driven by the few guests for whom the voucher could have been distortionary. This indicates that guests are influenced by the label attached to the voucher and thus spend too much on the targeted good, violating fungibility.

The laboratory experiment offers an even more controlled and well-defined setup. Again, subjects could consume two goods and had at their disposal a cash budget and either an in-kind grant or a cash grant. We induced a payoff function by specifying monetary payoffs for all possible consumption levels. We thus know the optimal decision and can guarantee that the in-kind grant was non-distortionary for every subject. We find that, as in the field experiment, subjects spend significantly more on the targeted good when the grant is given in kind. They also choose consumption bundles further away from the optimal decision and thus overall earn less money than subjects who receive the cash grant. A major advantage of the laboratory experiment is that we can collect further information about subjects to investigate the mechanisms underlying the treatment effect. Since using heuristics like narrow bracketing or mental accounting greatly reduces the complexity of the consumption decision, subjects who have lower cognitive ability should be more prone to this kind of cognitive bias. This hypothesis is confirmed by the data: the treatment difference is driven by subjects with lower cognitive skills. Moreover, subjects' tendency to violate fungibility is not influenced by their inclination to honor the stated preferences of a potential benefactor. This result thus points to narrow bracketing or mental accounting as the underlying reason for the violation of fungibility in the lab experiment.[2]

Our paper provides several novel insights: First, while there is a growing number of papers showing that investment decisions are often not in line with fungibility, we reveal a violation of fungibility for consumption decisions. For example, Choi et al. (2009) show that investors do not consider their existing portfolio when deciding how

[2]Frederick (2005), Benjamin et al. (2006), and Casari et al. (2007) also find that people with lower cognitive skills tend to act in accordance with theories of boundedly rational behavior whereas people with higher cognitive skills are more likely to behave in line with standard economic theory.

to invest their 401(k) contribution (see also Odean 1998). Second, and related to the first point, we demonstrate that many subjects violate fungibility in a setting where risk aversion or loss aversion cannot play a role, since they face a riskless choice in which no losses are possible. Most earlier work focused on risky settings and relied on risk or loss aversion for their predictions (e.g., Tversky & Kahneman 1981, Gneezy & Potters 1997, Thaler et al. 1997, Rabin & Weizsäcker 2009). Third, previous laboratory experimental studies on narrow bracketing of consumption decisions were not incentivized (e.g., Heath & Soll 1996, O'Curry 1997). Since subjects might use simplifying heuristics more readily if they do not face a payoff penalty for suboptimal decisions, these studies might have overestimated the prevalence of such heuristics. By using an incentivized laboratory experiment, we exclude this possibility. Fourth, the only existing studies investigating consumption decisions in incentivized environments analyze how people spend a gift or a windfall gain (e.g., Bodkin 1959, Arkes et al. 1994, Milkman & Beshears 2009). Most windfall gains are negligibly small compared to life-time wealth and should not alter spending behavior if customers treat wealth and windfall gain as fungible; but these studies find that people spend more after receiving a gift or windfall gain. If, however, the receipt of a gift or windfall gain per-se changes consumption patterns (e.g., because of a change in the recipient's mood[3] or for other reasons), it is not evident whether a change in spending can be clearly linked to a violation of fungibility. This cannot impact the main treatment comparisons in our study as subjects in both treatments receive a voucher of identical amount and only the type of voucher differs. Finally, by replicating the setup of the field experiment under laboratory conditions, we demonstrate the usefulness of laboratory experiments in complementing field evidence. In the laboratory, we can elicit extensive information about the subjects and about their decision processes. This allows us to explore the underlying mechanisms driving behavior and makes it possible to understand and interpret the results from the field (see also Falk & Heckman 2009, Charness & Villeval 2009).

It is important to investigate whether individual decisions are in line with fungibility to better understand consumer behavior in general. But the specific design of our study also suggests implications for public policy. We show that many consumers indeed do not treat money as fungible. Therefore, in-kind benefits will distort consumers' decisions more strongly than previously thought. Our results suggest, for instance, that part of the rent increase induced by housing benefits is due to a violation of fungibility. In our view, this problem can be mitigated by

[3]For some examples of the extensive research in psychology and marketing on the influence of mood on (consumption) choices, see Kahn & Isen (1993), Lewinsohn & Mano (1993), Groenland & Schoormans (1994), Winkielman et al. (2005), or Qiu & Yeung (2007).

linking housing benefits less saliently to rent payments to make it easier for tenants to treat this targeted grant as fungible. The periodicity of the benefit payments, for instance, could be chosen to differ from the periodicity of the rent payments. There are, however, other applications for which it might be desirable if recipients violate fungibility. If the government strongly believes that the consumption decisions of some households are not optimal (e.g., too little spending on child-related goods), it could use a violation of fungibility to improve these consumption decisions. By simply stating the intended use of the grant or by giving an in-kind grant that is non-distortionary, consumers who violate fungibility could be induced to buy more of the targeted good. This would not restrict or influence the consumption decisions of rational households, in line with the idea of libertarian paternalism proposed by Thaler & Sunstein (2003).

The paper is organized as follows: The general design of both experiments is described in Section 2. Section 3 reports the detailed design and results of the field experiment. Section 4 presents results of the laboratory experiment. Section 5 concludes.

2 Experimental Design

Our two experiments are designed to create tightly controlled environments in which we can directly test whether consumers treat money as fungible. We examine this question in a simple two-goods consumption case by investigating how consumers spend different kinds of lump-sum grants. Assume that a consumer has a cash budget of amount C at his disposal and additionally receives a grant of amount G. In the *Cash treatment*, the grant is given lump-sum in cash. In the *Label treatment*, the grant has the same amount but it is an in-kind grant, i.e., it has to be spent on one of the two goods, the targeted good. To illustrate, consider the indifference curve diagram in Figure 1. The targeted good (t) is on the horizontal axis and the other good (o) is on the vertical axis. For simplicity, the price of the targeted good is normalized to $p_t = 1$. The dot-and-dash line is the budget constraint if the consumer has only the cash budget C at his disposal. Assume that the optimal consumption bundle for this budget constraint is A'. The dashed line is the budget constraint in the Cash treatment, given by the sum of C and G. The optimal consumption bundle for this budget constraint is A. In the Label treatment, the grant is paid in kind; the consumer faces a kinked budget constraint (solid line). The crucial feature of our design is that the amount of the grant G is lower than the amount t^A spent optimally on the targeted good. Thus, the consumer can reallocate parts of his cash budget to still reach the first-best choice A. The in-kind grant in the Label treatment is therefore *non-distortionary*. Under the assumption that subjects treat money as fungible, consumption should be identical across treatments; treatments merely differ in the label attached to the grant.

Now consider a consumer who does not treat money as fungible. In the Cash treatment, we would still expect such a consumer to choose the first-best bundle A as both income components are cash. The difference to the standard model occurs in the Label treatment. The consumer will allocate the cash endowment optimally (bundle A'). The grant will, however, be spent disproportionally on the targeted good as the consumer does not take advantage of the possibility to reallocate parts of his cash budget. In the case of complete non-fungibility, this results in a consumption of bundle B where $t^B = t^{A'} + G$ (see Figure 1). If both goods are normal, $t^B \geq t^A$.[4]

Therefore, if at least some subjects do not treat money as fungible, we should expect average consumption of the targeted good in the Label treatment to be higher than in the Cash treatment. We cannot tell directly from this setup which underlying mechanism causes a potential violation of fungibility. In the lab experiment, however, we are able to collect further information on the subjects and their decision process that will shed light on the possible channels.

In the next section, we present the detailed design and results of the field experiment. The lab experiment is discussed in Section 4

3 Field experiment

3.1 Design of the Field Experiment

We chose a restaurant as setting to conduct the field experiment as it is the ideal environment to implement the setup we want to investigate: since guests typically consume a minimal amount of two distinct goods (beverages and meals), giving an in-kind grant that is smaller than this minimal amount will not distort the consumption decision. The experiment took place in a wine restaurant situated in the palatinate, a well-known wine-growing region of southern Germany. The restaurant itself is located in a winery. Usual per-person spending in this restaurant

[4]If a violation of fungibility is driven by narrow bracketing or mental accounting, this reasoning depends on the order in which cash budget and grant are spent. If the consumer spent the grant first, he would be able to allocate the cash budget so as to reach bundle A. In the experiments, we are therefore testing the joint hypothesis of fungibility and order of spending.

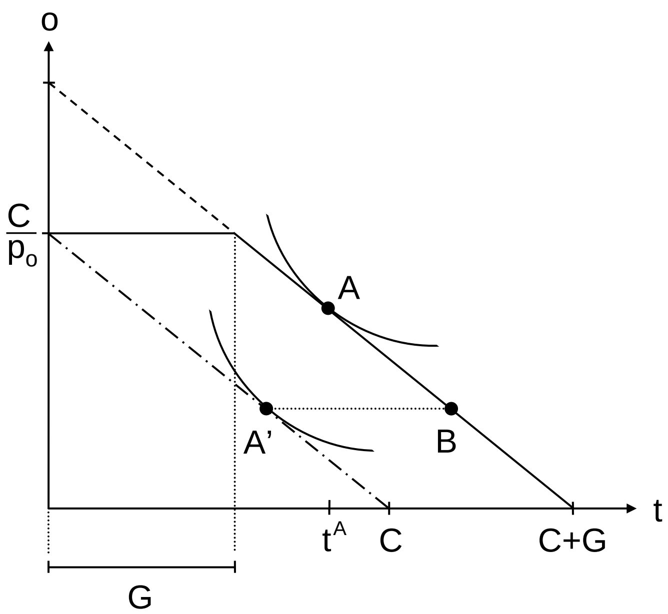


Figure 1: *The targeted good (t) is on the horizontal axis, the other good (o) is on the vertical axis. The dashed line is the budget constraint when the grant is given in cash. The solid line is the budget constraint when the grant is given in kind. The dot-and-dash line is the budget constraint when no grant is given.*

is about 40 euros (∼54 USD at the time of the experiment); about 40 percent of the total is spent on beverages. This setting thus matches the two-goods case presented in Figure 1 particularly well.

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 person (∼11 USD). The type of voucher differed by day: on days of the *Cash treatment*, vouchers were given as "gourmet voucher" that could be spent on both beverages and meals. This treatment serves as our primary control treatment. On days of the *Label treatment*, vouchers were given as a "gourmet beverage voucher". These vouchers were restricted to be spent on beverages. We knew from communication with the owner of the restaurant that, without getting a voucher, the overwhelming majority of guests consumes beverages worth more than 8 euros (it is very unusual to not consume beverages in German restaurants; water must also be purchased). Therefore, the beverage voucher should be non-distortionary.[5] Both types of vouchers had to

[5] The design of the experiment made it impossible to rule out that some participants could have wanted to consume less than the amount of the grant, as total consumption had a high variance. In Section 3.2 we will present results that show that our treatment effects are not driven by these participants. In the laboratory experiment, described in the next section, we can ensure that

be redeemed the same evening. In a third treatment, the *Baseline treatment*, guests did not receive any voucher.

We have data on 552 guests. Overall, 107 vouchers were distributed in the Label treatment and 89 vouchers in the Cash treatment, one per person.[6] 356 persons participated in the Baseline treatment. We consider each table in the restaurant as one independent observation and calculate all values per person. Since we distributed one voucher per person, we can relate per-person consumption directly to the amount of a single voucher. This leaves us with 37 independent observations in the Label treatment, 34 in the Cash treatment, and 116 in the Baseline treatment. During the observed period, 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; we also know how many persons correspond to each bill.

3.2 Results of the Field Experiment

First, we demonstrate that consumption of the targeted good (beverages) is higher in the Label treatment than in the Cash treatment.

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

We document consumption averages for the three treatments in Table 1. Participants in the Label treatment—who receive a beverage voucher—spend on average 18.94 euros per person on alcoholic and non-alcoholic beverages, 3.90 euros more than participants in the Cash treatment and also more than participants who didn't receive either voucher. This treatment difference is very large compared the value of the grant (8 euros) and to average beverage consumption across treatments (16.21 euros) and in line with the hypothesis that consumers violate fungibility. Participants in the Cash treatment spend more on meals than subjects in both other treatments. This translates into higher total consumption in both voucher treatments compared to the Baseline treatment.

To test whether these differences are statistically significant, we use OLS regres-

---

the grant is non-distortionary for all subjects by choosing an appropriate payoff function and by endowing every subject with the same budget.

[6]The restaurant first issued all beverage vouchers. From the next day on, the remaining vouchers were issued as bill vouchers. This was done for practical reasons, as the restaurant feared that switching treatments on a daily basis would be too confusing for the waiters (see, e.g., Bandiera et al. (2005) for a field experiment with similar treatment sequencing).

Table 1: Average Consumption Across Treatments

	Baseline treatment	Cash treatment	Label treatment
Beverage consumption	15.69	15.03	18.94
Meal consumption	24.18	27.68	25.86
Total consumption	39.87	42.71	44.80

Notes: All amounts denoted in euro.

sions and regress per-person consumption on a dummy for receiving a voucher at all and a dummy for being in the Label treatment. Our main focus will be on the coefficient of the Label dummy, since the comparison between Label and Cash treatment allows for the cleanest interpretation whether fungibility is violated. Participants in both of these two treatments receive a voucher and changes in consumption patterns because of receiving a voucher per-se cannot influence the treatment comparison.

Table 2 reports results of regressions with per-person beverage consumption as dependent variable. We find that receiving a voucher per-se has no significant impact on beverage consumption. But receiving a beverage voucher instead of a bill voucher significantly increases spending on beverages (column 1). This means that by merely attaching a label to the grant consumption of the targeted good is significantly increased compared to an unlabeled voucher. Since beverage and meal consumption are usually regarded as complements and since they are indeed highly correlated in our sample ($r = 0.442$), we control for meal consumption in column 2 to isolate the direct effect of the type of voucher on beverage consumption. Also, one might interpret meal consumption as a proxy for an important determinant of (beverage) consumption: total wealth. Richer people arguably spend more on beverages and also more on meals. By controlling for total wealth the precision of our treatment effect estimation should thus improve. Indeed, we find that the p-value of the treatment effect is now lower (column 2). The point estimate of the treatment effect (Label vs. Cash) increases a little bit since meal consumption is higher in the Cash treatment than in the Label treatment. This is also the reason why the voucher dummy has as weakly significant negative coefficient when we control for meal consumption. The treatment effect stays significant if we additionally include outside temperature and dummies for the day of the week as control variables (column 3).

While subjects in the Label treatment use the voucher to increase their beverage consumption, subjects in the Cash treatment use their voucher predominantly to

Table 2: Treatment Effect on Beverage Consumption

Dependent variable: Beverage consumption (in euro per person)

	Full sample			Restricted sample		
	(1)	(2)	(3)	(4)	(5)	(6)
1 if Either voucher	-0.652	-2.200*	-1.521	0.036	-1.054	-0.495
	(1.339)	(1.151)	(1.394)	(1.590)	(1.513)	(1.732)
1 if Label treatment	3.903**	4.707***	4.512**	3.398**	4.210**	3.748**
	(1.855)	(1.638)	(1.599)	(1.608)	(1.608)	(1.656)
Meal consumption		0.443***	0.448***		0.319***	0.326***
		(0.069)	(0.069)		(0.084)	(0.083)
Outside temperature (in °C)			-0.197			-0.157
			(0.140)			(0.173)
Controls for day of the week	No	No	Yes	No	No	Yes
Constant	15.687***	4.971***	8.387***	17.384***	9.176***	13.308***
	(0.626)	(1.584)	(2.470)	(0.832)	(2.211)	(3.261)
N.Obs.	187	187	187	157	157	157

Notes: OLS estimates. Columns 1-3 report results for the full sample; columns 4-6 report results for a sample from which potentially distorted participants are excluded (see text for details). Robust standard errors clustered on day are in parentheses. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Table 3: Treatment Effect on Meal Consumption

Dependent variable: Meal consumption (in euro per person)

	Full sample		
	(1)	(2)	(3)
1 if Either voucher	1.679**	0.176	0.209
	(0.698)	(0.679)	(0.961)
1 if Cash treatment	1.814***	3.619***	3.349***
	(0.597)	(0.529)	(0.741)
Beverage consumption		0.462***	0.462***
		(0.058)	(0.057)
Outside temperature (in °C)			0.016
			(0.107)
Controls for day of the week	No	No	Yes
Constant	24.184***	16.929***	15.894***
	(0.599)	(1.168)	(1.889)
N.Obs.	187	187	187

Notes: OLS estimates. Robust standard errors clustered on day are in parentheses. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

increase their meal consumption. Table 3 reports estimates showing that participants in the Cash Treatment spend significantly more on meals than participants in the Label treatment (column 1). This effect is robust to controlling for beverage consumption, outside temperature, and day of the week (columns 2 and 3). Receiving either voucher also increases meal consumption compared to participants in the Baseline treatment, but the direct effect of receiving either voucher on meal consumption goes away once we control for the indirect effect via beverage consumption.

Because of the high variance in total per-person consumption, we could not exclude the possibility that some guests would initially have liked to spend less on beverages than the amount of the voucher. For these guests, the voucher was distortionary and they might have increased their beverage consumption to use the voucher in full and not because of a violation of fungibility. Indeed, in 16 percent of observations in the Cash and Baseline treatments, absolute beverage consumption is lower than the value of the voucher. To take account of these potentially distorted decisions, we sort all observations by per-person beverage consumption and exclude the lowest 16 percent in each treatment. As treatment assignment is random, the distribution of initial willingness to consume beverages should be equal across treatments and the grant should therefore be non-distortionary for all participants in the restricted sample. In columns 4–6 of Table 2 we repeat the estimations of columns 1–3 for the restricted sample. Results do not change. The treatment effect between Label and Cash treatment is of very similar size and still significant even though the number of observation is smaller. Only the intercept is larger, which is not surprising as we exclude the low-consumption observations for these regressions. We can thus rule out that the treatment effect is driven by the minority of participants for whom the voucher could have been distortionary.

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 in beverage consumption. As a further robustness check, we test for such a trend with the data we collected before and after the two main treatments when guests did not receive either voucher. If a time trend existed in the two main treatments, 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 than before but this difference is not significant.[7]

Next, we analyze the additional spending on beverages in the Label treatment in more detail.

> **Result 2:** *Participants in the Label treatment do not consume more beverages in terms of volume but they consume more expensive beverages.*

To see whether participants in the Label treatment consume a larger volume of beverages, we regress the volume of consumed beverages on the same set of explanatory variables as in Table 2 (see Table A.1, columns 1–3, in the appendix). The effect of the voucher and of the type of voucher on consumed volume is very small (e.g., Cash vs Label: 0.050 ltr, or 2 fl oz, in column 1) and never significant at the 5 percent level. As in Table 2, we also report results for the sample restricted to participants for whom the grant is non-distortionary (columns 4–6). We again find that the volume of consumed beverages does not differ between treatments.

In contrast, the type of voucher strongly influences the average price of the

[7]Guests spend on average 15.38 euros on beverages before the two main treatments and 16.27 euros after. When we repeat the estimations of Table 2, columns 1-3, with an additional dummy for observations before the two main treatments the p-values of the dummy are $p = 0.494$, $p = 0.889$ and $p = 0.302$.

consumed beverages. Receiving any voucher at all does not have a significant impact. But if the voucher is targeted to beverages, the average price rises sharply and significantly. Participants in the Label treatment spend on average 21.91 euros per liter, 3.52 euros more than participants in the Cash treatment (see Table A.2 in the appendix). We therefore conclude that spending on beverages is higher in the Label treatment because participants use the voucher to buy more expensive beverages.

So far we have argued that—if guests treated money as fungible—receiving a beverage voucher compared to a bill voucher should not alter consumption behavior. The same argument can also be applied to the comparison of guests who receive either voucher and guests without voucher. The 8-euro increase in lifetime income (by receiving the voucher) can surely be neglected. According to standard consumer theory, we should thus expect consumption not to be influenced by receiving a voucher compared to not receiving a voucher. If fungibility is violated, receiving any voucher could, however, influence consumption. Keep in mind, though, that the treatment comparison between voucher treatments and the Baseline treatment might also be influenced by several effects unrelated to fungibility. For example, it might be that receiving a voucher as gift makes the recipient spend more just because they get into a different mood by receiving a gift (see, e.g., Lewinsohn & Mano 1993, Winkielman et al. 2005). Mood changes might also lead to *less* spending, or spending could be reduced for other reasons, for example because receiving a voucher makes it more salient that the meal has to be paid for at the end. We thus report this comparison only for completeness. The treatment comparison Label vs. Cash allows for cleaner conclusions as to whether participants decide in line with fungibility or not since participants receive a voucher in both treatments.

> **Result 3:** *Overall spending is higher in both voucher treatments compared to the Baseline treatment.*

As shown above, beverage consumption is significantly higher in the Label treatment and meal consumption is higher in the Cash treatment. While beverage or meal consumption individually are not robustly different between the two voucher treatments and the Baseline treatment (see Tables 2 and 3), total consumption does differ. In Table 4, columns 1 and 2, we regress total consumption on a dummy for receiving either voucher and the same controls as in Table 2. Participants in the Baseline treatment spend on average 39.87 euros per head. Participants in the two voucher treatments spend significantly more, on average 43.80 euros. In columns 3 and 4, we repeat these regressions for the restricted sample of participants for whom the beverage voucher should be non-distortionary. The point estimates remain high

Table 4: Treatment Effect on Total Consumption

Dependent variable: Total consumption (in euro per person)

	Full sample		Restricted sample	
	(1)	(2)	(3)	(4)
1 if Either voucher	3.930**	4.918**	3.900**	3.792
	(1.629)	(2.167)	(1.572)	(2.367)
Outside temperature (in °C)		-0.348		-0.117
		(0.294)		(0.347)
Controls for day of the week	No	Yes	No	Yes
Constant	39.871***	44.883***	43.100***	46.489***
	(1.012)	(5.022)	(1.277)	(6.599)
N.Obs.	187	187	157	157

Notes: OLS estimates. Columns 1 and 2 report results for the full sample; columns 3 and 4 report results for a sample from which potentially distorted participants are excluded (see text for details). Robust standard errors clustered on day are in parentheses. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

but the effect is only significant in the regression of column 3.[8]

4 Laboratory experiment

Combining a field and a laboratory experiment has many advantages. While we can analyze a more natural setting in the field experiment where participants don't know that they take part in an experiment, the laboratory offers three complementary features. First, the lab allows for a more tightly controlled environment. We can keep the budget equal for all subjects which is not possible in the field. And since we know the optimal decision in the lab (see below for details) we can guarantee that the grant is non-distortionary for all subjects and thus do not have to exclude potentially distorted subjects. Second, we can gather a more informative measure of behavior than in the field: subjects decide on consumption for two budget constraints

[8]Total consumption is not significantly different between Label and Cash treatment. If we include an additional dummy for the Label treatment in the regressions of Table 4, p-values of the Label dummy range from 0.298 to 0.815.

which allows calculating a within-person measure of behavior. Finally, and most importantly, we have more information about the subjects in the lab. We can therefore investigate determinants of subjects' behavior by linking their consumption decisions to individual subject characteristics and to additional features of their behavior, like the time needed to reach a decision.

4.1 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 cash budget that they could spend on two goods. For each good, we defined a payoff function by specifying monetary payoffs for all possible consumption levels. A subject's total payoff was the sum of the payoffs for each of the two goods in both stages. In the first decision stage, which we will call *baseline stage*, subjects received a cash budget of 50 money units which they could allocate freely on the two goods. The baseline stage was identical in both treatments. The second stage, called *grant stage*, is our main treatment stage. In the grant stage, subjects again had an endowment of 50 money units at their disposal and additionally received a grant of 30 money units. The only difference between the two treatments was the type of the grant. In the *Cash treatment*, the grant was given as an unconditional cash grant. In the *Label treatment*, the grant was given as an in-kind grant, i.e., the money had to be spent entirely on the targeted good. Parameters were chosen such that the in-kind grant was by design non-distortionary for all subjects. By shifting the remainder of their budget appropriately, subjects could reach the same optimal consumption bundle in both treatments. For a rational subject, the only treatment difference was therefore the label attached to the grant.

The exact specification of the payoff functions is presented in Table 5. For each good, payoff increases in consumption and marginal payoff weakly decreases. Prices per unit were $p_t = 3$ for the targeted 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 grant is worth 10 units of the targeted good and the consumption bundles (t, o) displayed in Figure 1 are as follows: the optimal consumption bundle in the baseline stage is $A' = (12, 7)$; the optimal bundle in the grant stage is $A = (13, 20)$; the bundle B is $(22, 7)$.[9]

[9]Obviously, when we say "optimal" decision we mean the "financially optimal" decision. It might well be that for subjects with lower cognitive skills it is overall not optimal to invest the

Consumed units	0	1	2	3	4	5	6	7	8	9	10	11	12
Payoff													
Targeted 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													
Targeted 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 5: *Payoff functions in the laboratory experiment. "Targeted good" denotes the good that the grant had to be spent on in the second stage of the Label treatment.*

In order to make the difference between the initial endowment and the grant 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 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.[10] 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 set of control questions to ensure that they understood the task (see Appendix B for an English translation of the instructions). Detailed instructions for the second stage were given only 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

additional effort to reach a higher monetary payoff, see Section 4.2.2 for details.

[10]For a similar real-effort task, see Abeler et al. (forthcoming). 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. This was the case for 7 subjects who will be excluded from the analysis.

using z-Tree and ORSEE (Fischbacher 2007, Greiner 2004). 150 subjects completed the real-effort task, 75 in each treatment. Payoff points (cf. Table 5) were paid out after the experimenter 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 euros. On average, subjects earned 12.20 euros (∼14.80 USD at the time of the experiment). Sessions lasted between 60 and 70 minutes.

4.2 Results of the Laboratory Experiment

First, we show that, as in the field experiment, giving a labeled grant instead of a cash grant increases consumption of the targeted good. Then we explore possible determinants for the observed violations of fungibility.

4.2.1 Consumption in the Experiment

Before we turn to the main stage of the experiment, the grant stage, we analyze consumption decisions in the baseline stage. The design of the baseline stage was the same in both treatments. Accordingly, we find that behavior in this stage is not different across treatments. In Table 6, column 1, we regress consumption of the (later to be) targeted good in the baseline stage on a dummy for the Label treatment.[11] We use tobit estimates to account for the fact that subjects could only buy up to 25 units of the targeted good.[12] In column 2, we also control for gender, age, and major of subjects. In both specifications, the treatment effect is very small and not significant. This means that our random assignment to treatments worked. The regressions further show that average consumption (11.4 units) is close to the optimum of 12 units. Subjects apparently have no problem understanding the decision problem and take the decision seriously. This is confirmed when we take the absolute distance to the optimal consumption level as dependent variable; this measure also treats too low consumption as error. Again, treatments are almost indistinguishable (see Table 7, columns 1 and 2) and most subjects choose consumption levels close to the optimum (average distance is 1.5 units). We are therefore confident that the

[11]For ease of exposition, we report only the consumption of the targeted good. Consumption of the other good can then be readily calculated as very few subjects in the experiment choose a consumption bundle that is not on the budget frontier (13 out of 300 decisions). Our results do not change if we exclude subjects who choose bundles that are not on the budget frontier.

[12]5 out of 150 subjects buy the maximum of 25 units in the grant stage. All results in this section also hold if we employ OLS regressions instead of tobit.

experimental setup allows for meaningful interpretation and that the experimental incentives work.

Next, we analyze outcomes in the grant stage.

> **Result 4:** *Consumption of the targeted good is significantly higher in the Label treatment.*

In the grant stage, subjects in the Label treatment buy too much of the targeted good. They buy 16.6 units on average, compared to 14.4 units in the Cash treatment and an optimal consumption level of 13. The estimates in Table 6, columns 3 and 4, show that the treatment effect is highly significant and remains unaffected when we control for subjects' age, gender, and major. In column 4, we also control for the consumption of the targeted good in the baseline stage as this might influence consumption in the grant stage due to inertia or anchoring. The treatment effect is also significant when we take the distance to the optimal consumption as dependent variable (Table 7, column 3). In column 4 of Table 7, we additionally control for the distance to the optimal consumption level in the baseline stage, taking this as a proxy for how well subjects are able to deal with the general decision problem at hand. Again, subjects in the Label treatment choose consumption bundles significantly further away from the optimal bundle. By consuming too much of the targeted good, subjects in the Label treatment leave money on the table as their choices translate into significantly lower payoffs.[13]

These results confirm the main finding of our field experiment: even in this stylized and tightly controlled environment subjects do not treat money as fungible. Our next result documents the considerable heterogeneity we observe in behavior across individuals.

> **Result 5:** *The treatment difference is to a large part caused by subjects who increase their consumption by the full amount of the grant.*

The two-stage design of our experiment enables us to compute an intra-person measure of behavior by comparing decisions in the grant stage to decisions in the baseline stage. A histogram of the intra-person change in consumption is shown in Figure 2. The grant was worth 10 units of the targeted good. In line with the results reported above, one can see clearly that the consumption increase is higher in the Label than in the Cash treatment (t-test, $p < 0.001$). What is more interesting is that decisions

[13]With payoff in the grant stage as dependent variable, the p-values of the treatment dummy are 0.014 and 0.016 in tobit regressions equivalent to the specifications in Table 6, columns 3 and 4 (limiting payoff at the maximal reachable level).

Table 6: Absolute Consumption in the Laboratory Experiment

	Dependent variable: Consumption of the targeted good (in units)					
	Baseline stage		Grant stage		Grant stage Restricted sample	
	(1)	(2)	(3)	(4)	(5)	(6)
1 if Label treatment	-0.133	-0.124	2.205***	2.145***	1.237*	1.308**
	(0.350)	(0.354)	(0.636)	(0.582)	(0.649)	(0.579)
Consumption in baseline stage				0.607***		0.690***
				(0.135)		(0.136)
Controls for gender, age, major	No	Yes	No	Yes	No	Yes
Constant	11.440***	12.150***	14.442***	7.679***	14.366***	5.625*
	(0.248)	(1.417)	(0.449)	(2.848)	(0.432)	(2.963)
N.Obs.	150	150	150	150	133	133

Notes: Tobit estimates (lower limit 0, upper limit 25 units). The dependent variable is consumption of the targeted good in the baseline stage (columns 1–2) or in the grant stage (columns 3 to 6). Columns 1–4 report results for the full sample; columns 5 and 6 report results for subjects who did *not* increase their consumption of the targeted good by exactly 10 units. Standard errors are in parentheses. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Table 7: Distance to Optimal Consumption

	Dependent variable: Absolute distance to optimal consumption of the targeted good (in units)					
	Baseline stage		Grant stage		Grant stage Restricted sample	
	(1)	(2)	(3)	(4)	(5)	(6)
1 if Label treatment	-0.000	0.031	2.634***	2.274***	1.598*	1.674**
	(0.455)	(0.456)	(0.767)	(0.675)	(0.815)	(0.687)
Distance to optimum in base. stage				0.811***		0.972***
				(0.205)		(0.213)
Controls for gender, age, major	No	Yes	No	Yes	No	Yes
Constant	0.741**	2.496	1.020*	-1.350	0.941*	-2.230
	(0.333)	(1.855)	(0.558)	(2.692)	(0.560)	(2.769)
N.Obs.	150	150	150	150	133	133

Notes: Tobit estimates (lower limit 0, upper limit 12 units). The dependent variable is the absolute distance to optimal consumption of the targeted good in the baseline stage (columns 1–2) or in the grant stage (columns 3 to 6). Columns 1–4 report results for the full sample; columns 5 and 6 report results for subjects who did *not* increase their consumption of the targeted good by exactly 10 units. Standard errors are in parentheses. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

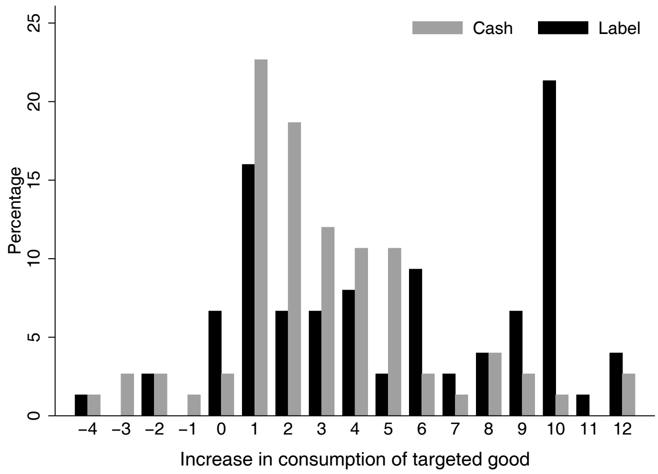


Figure 2: *Consumption increase of the targeted good from baseline stage to grant stage. The grant is worth 10 units of the targeted good.*

are highly heterogeneous in the Label treatment. The most frequent consumption increase in the Cash treatment is by either 1 or 2 units, often leading to a choice of the optimal consumption bundle in the grant stage. In contrast, the modal choice in the Label treatment is a consumption increase by 10 units, i.e., subjects spending the entire grant on the targeted good on top of the consumption from the baseline stage. Subjects who treat income sources as completely non-fungible will do exactly this (cf. bundle B in Figure 1). In the Label treatment, 21 percent of subjects spend the whole grant on the targeted good, while this is true for only 1 percent of subjects in the Cash treatment.[14] These subjects drive a large part 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 and less significant (see Table 6, columns 5 and 6). The same is true if we take distance to optimal consumption as dependent variable (Table 7, columns 5 and 6).

We thus find that not all subjects are equally likely to violate fungibility. In the the next section, we explore what determines this heterogeneity in behavior.

[14] As a consequence of their consumption decision, subjects who spend the entire grant on the targeted good earn less than all other subjects and also less than the other subjects in the Label treatment. If we regress payoff in the grant stage on a dummy for subjects who increase consumption by exactly 10 and control for treatment, the dummy's p-value is $p = 0.005$. If we restrict the sample to the Label treatment the p-value is $p = 0.011$.

4.2.2 Determinants of Behavior

Several behavioral theories predict a violation of fungibility, most prominently theories of cognitive biases like narrow bracketing or mental accounting. However, consumers' spending of non-distortionary in-kind grants might also violate fungibility because of a perceived obligation to reciprocate the receipt of the grant by complying with the stated preferences of the giver. We cannot say how important this mechanism is for recipients of state benefits or for the participants in our field experiment but it should be diminished in our highly stylized lab experiment. Consumption decisions were about abstract goods and the payoffs for each good were converted into real money directly after the experiment. In addition, the instructions did not state any kind of preference over how the money should be allocated across the two goods.[15] The fact that many subjects increase their consumption by exactly the amount of the grant also speaks against a stated-preference effect: Even if an otherwise rational subject wants to honor the stated preferences of the grant giver, consumption could be increased by more or less than the amount of the grant. Spending exactly this amount is only predicted by narrow bracketing or mental accounting.[16]

But can we find more direct evidence that narrow bracketing or similar cognitive biases drive the treatment effect in the lab? A consumer who brackets his decisions narrowly will violate fungibility but also greatly reduce the complexity of the consumption decision. Subjects who have difficulties with abstract reasoning and complex decisions 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. In their survey of narrow bracketing, Read et al. (1999) also conjecture that: "Cognitive limitations—in perception, attention, memory, and analytical processing, etc.—are one important determinant of bracketing." However, there is so far no evidence for this conjecture. If narrow bracketing drives behavior in our experiment, we should find that subjects' cognitive skills are negatively correlated with the treatment effect.

[15] Any suggestion (or informed-principal) effect should also be reduced in the lab experiment. Such an effect assumes that subjects believe that the benevolent grant provider has more information about the optimal consumption level than they do and that the label attached to the grant is informative about optimal consumption. Thus, subjects rationally follow the suggestion of the label. But in our lab experiment, subjects do have all necessary information available and could easily check whether the suggestion of the label leads to higher payoffs, which was indeed *not* the case.

[16] We will also test for the importance of perceived obligation directly, see below.

Our next result supports this hypothesis.

> **Result 6:** *The treatment difference in consumption is driven by subjects with lower cognitive skills.*

We use subjects' math grade in their final high school exam as a proxy for their cognitive, especially nonverbal, ability. This is a good proxy for several reasons. Math is a compulsory course that every high school pupil has to take; the grade is highly incentivized since it is used to determine university entrance and employment decisions; the grade covers performance over a long period (usually 2 years), reducing measurement error; and, most importantly, it is highly correlated with other measures of intelligence and cognitive ability.[17] The grades were elicited in the post-experimental questionnaire. Grades range from 1 (best grade) to 5 (fail), a higher grade thus indicates a poorer performance.

In Table 8, columns 1–4, we regress consumption of the targeted good on a treatment dummy, the math grade of subjects and an interaction of grade and treatment. We include the same control variables as in Table 6. By including dummies for subjects' major we control for any additional effect of university (math) education that might influence decision making. The specification in column 1 of Table 8 shows that math grade has *no* effect in the baseline stage of either treatment. Thus, the math grade does not just capture being better able to tackle the consumption decision posed in the experiment. Also in the grant stage (column 2), there is no effect on behavior in the Cash treatment (the coefficient of grade is very small and not significant). Only the effect on behavior in the Label treatment is pronounced and significant. If the math grade gets worse by one grade, a subject increases consumption by 1.01 units on average. In fact, the math grade captures the whole treatment effect, the treatment dummy is not significant anymore. The result that the math grade has only an effect on behavior in the Label treatment but not in the Cash treatment is corroborated when we estimate separate regressions for each treatment (columns 3 and 4). This avoids identification of the impact of the math grade by imposing identical coefficients on the control variables. While the effect in

[17] For example, Deary et al. (2007) found in a large, representative sample that the correlation between an individual's general intelligence factor g at age 11 and their math grade at age 16 was 0.77. g is what standard cognitive-ability tests (or "IQ"-tests) try to measure. This correlation was higher than the correlation of g with the grade of any other course. An alternative measure of cognitive ability used by other studies is the subject's SAT or ACT score (e.g., Benjamin et al. 2006, Casari et al. 2007). The correlation between SAT-score and g has been estimated as 0.70 (Brodnick & Ree 1995) or 0.82 (Frey & Detterman 2004), very similar to the correlation between math grade and g.

the Cash treatment is not significant, it is again sizable and significant in the Label treatment. The math grade coefficients are significantly different between these two specifications ($\chi^2(1)$ test, $p = 0.008$).[18]

All these results hold if we take absolute distance to the optimal consumption as dependent variable (see Table 8, columns 5–8). In these regressions, we control for the distance to optimal consumption in the baseline treatment. This should be a good proxy for how well subjects can deal with the payoff functions and how serious they take the decision. In these specifications we can therefore further isolate the effect of cognitive skills on the treatment difference. These regressions yield the same results: math grade does not influence behavior in the baseline stage or in the grant stage of the Cash treatment. But the worse the math grade, the larger is the distance to the optimal consumption in the Label treatment.[19]

We have shown before that subjects who increase consumption by 10 units, i.e., who spend the entire grant on the targeted good, account for a large part of the treatment effect. This behavior is also correlated with subjects' math grade: The worse the math grade, the more likely it is that a subject increases consumption by exactly 10 units.[20] Interestingly, subjects who spend the entire grant on the targeted good also differ in another decision characteristic: they decide much faster than the remaining subjects. They need on average 116 sec for their decision, whereas the other subjects need 267 sec, more than twice as long (t-test, $p < 0.001$). This difference suggests that spending the grant fully on the targeted good is the result of a simple decision heuristic rather than extensive deliberations.

Although feelings of perceived obligation to comply with the stated preference of

[18] The 7 subjects (3 in the Cash and 4 in the Label treatment) who did not complete the real-effort task and who were excluded from the experiment have a lower average math grade than the remaining sample (3.00 vs. 2.34). Result 4 thus underestimates the main treatment effect as we exclude subjects who on average are more prone to the treatment manipulation.

[19] In the German high school system, there are two types of math course: intensive and basic course (*Leistungskurs* and *Grundkurs*). If we control for the type of subjects' high school course, results of Table 8 do not change and the math grade stays significant. The course dummy is never significantly different from zero. Apparently, the math course dummy does not add much information beyond math grade and university major.

[20] In a probit regression within the Label treatment of a dummy for "increasing consumption by exactly 10" on a constant and the math grade, the marginal effect of the grade is 0.100 ($p = 0.010$), i.e., if the grade gets worse by one, the probability to increase consumption by exactly 10 units increases by 10 percentage points. We restrict the sample to the Label treatment, as only 1 out of 17 subjects who increase consumption by 10 is in the Cash treatment and to avoid interaction terms (Ai & Norton 2003).

Table 8: Impact of Cognitive Ability

	Dependent variable: Consumption of the targeted good				Absolute distance to optimal consumption			
	Baseline	Grant	Grant stage		Baseline	Grant	Grant stage	
	Full sample		CT	LT	Full sample		CT	LT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 if Label treatment	0.270	-0.165			-0.139	-0.783		
	(0.767)	(1.236)			(0.987)	(1.436)		
Math grade	-0.032	-0.173	-0.459	1.040**	-0.287	-0.203	-0.393	1.209**
	(0.228)	(0.366)	(0.333)	(0.461)	(0.289)	(0.431)	(0.416)	(0.489)
Math grade*Label	-0.174	1.011**			0.070	1.324**		
	(0.296)	(0.477)			(0.382)	(0.551)		
Consumption in baseline stage		0.626***	0.593***	0.813***				
		(0.133)	(0.153)	(0.221)				
Distance to optimum in base. stage						0.853***	0.621**	1.033***
						(0.201)	(0.236)	(0.332)
Controls for gender, age, major	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	12.174***	7.923***	2.112	8.348*	2.974	-0.852	-5.372	0.243
	(1.479)	(2.879)	(3.767)	(4.308)	(1.920)	(2.725)	(3.637)	(3.824)
N.Obs.	150	150	75	75	150	150	75	75

Notes: Tobit estimates. The dependent variable is consumption of the targeted good in the baseline stage in column 1 or in the grant stage (columns 3–4) and the absolute distance to optimal consumption of the targeted good in the baseline stage (column 5) or in the grant stage (columns 6–8). Columns 1–2 and 5–6 report results for the full sample; columns 3 and 7 report results for subjects in the Cash treatment and columns 4 and 8 for subjects in the Label treatment. The lower limit for the tobit estimation is 0 in all specification; the upper limit is 25 for columns 1–4 and 12 for columns 5–8. The math grade coefficients are significantly different between the specifications in columns 3 and 4 ($\chi^2(1)$ test, $p = 0.008$) and between columns 7 and 8 ($\chi^2(1)$ test, $p = 0.013$). Standard errors are in parentheses. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

the grant giver should be reduced in our laboratory experiment, we still measured subjects' feelings of obligation to see whether it affected their behavior. In the post-experimental questionnaire, we described in a short vignette that a couple spent less money on their kids than what they received as child benefits.[21] Subjects then had to state how "justified" or "appropriate" (on a scale of 1 to 6) they found this behavior. Note that the fictitious couple faces a similar decision situation as subjects did in the experiment. In both situations, a rational decision maker should not be influenced by the label attached to the grant. But in contrast to the lab decision, the provider of child benefits (i.e., the government) has clearly stated preferences about the final allocation of consumption. This measure of perceived obligation does, however, not predict behavior in the experiment. If we run a regression similar to the one in column 2 of Table 8 but replace math grade by our measure of obligation, neither the direct effect nor the interaction is significant ($p = 0.267$ and $p = 0.767$).

This does not preclude the possibility that in other environments the felt obligation to comply with the stated preferences of the giver additionally influences behavior. In our lab experiment, however, cognitive biases seem to drive the treatment effect that we uncover, supporting theories of narrow bracketing and mental accounting.

5 Conclusion

In this paper we pursued a dual research strategy by combining a natural field experiment and an incentivized laboratory experiment to test whether consumers treat money as fungible. Both experiments yield the same result: many subjects do not act in line with fungibility. In the lab, where we have more background information about subjects, this effect is driven by subjects with lower cognitive, especially mathematical, skills. This points to cognitive biases like narrow bracketing or mental accounting as mechanism underlying a violation of fungibility.

We argued that fungibility plays an important role in a setting where it has until now not been considered: the effect of in-kind benefits on consumption and market prices. Empirical studies have shown that a rise in housing benefits has

[21] The full text of the vignette was: "Mr and Mrs Miller have two children (5 and 8 years old). They earn a total amount of 2000 euros per month, after taxes. Additionally, they receive 180 euros child benefit per child, i.e., a total of 360 euros per month. Usually, they spend about 300 euros 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)".

lead to pronounced rent increases (see, e.g., Susin 2002, Gibbons & Manning 2006, Fack 2006). In addition, Laferrère & Le Blanc (2004) show that controlling for apartment and neighborhood characteristics, landlords discriminate between non-assisted tenants and tenants who receive housing assistance, charging the latter group higher rents. Our results suggest that this effect is partly due to a violation of fungibility and could thus be mitigated by linking housing benefits less saliently to rent payments to make it easier for tenants to treat this income source as fungible. In other situations, it might, however, be desirable if recipients violate fungibility. In the spirit of libertarian paternalism (Thaler & Sunstein 2003), one could influence consumption decisions by simply stating the intended use of a grant or by replacing cash grants with non-distortionary in-kind grants.

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 lower cognitive skills, suggesting that this behavior is a mistake and not driven by preferences. Once the rational solution becomes obvious to subjects, e.g., by learning or by explanation, it could be that subjects change their decision and choose the optimal solution. Further experiments would be needed to test this hypothesis. Another possible next step would be to use the laboratory as a testbed for institutions that could help subjects to treat money as fungible (cf. Plott 1997, Roth 2002).

6 References

Abeler, J., Falk, A., Goette, L. & Huffman, D. (forthcoming), 'Reference points and effort provision', *American Economic Review* .

Ai, C. & Norton, E. C. (2003), 'Interaction terms in logit and probit models', *Economics Letters* **80**, 123–129.

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.

Bandiera, O., Barankay, I. & Rasul, I. (2005), 'Social Preferences and the Response to Incentives: Evidence from Personnel Data', *Quarterly Journal of Economics* **120**(3), 917–962.

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

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

Blow, L., Walker, I. & Zhu, Y. (2007), 'Who benefits from child benefit?', *Discussion Paper, University of Warwick* .

Bodkin, R. (1959), 'Windfall Income and Consumption', *American Economic Review* **49**(4), 602–614.

Brodnick, R. J. & Ree, M. J. (1995), 'A structural model of academic performance, socioeconomic status, and Spearman's *g*', *Educational and Psychological Measurement* **55**(4), 583–594.

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. & Villeval, M. C. (2009), 'Cooperation and competition in intergenerational experiments in the field and the laboratory', *American Economic Review* **99**(3), 956–978.

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

Choi, J. J., Laibson, D. & Madrian, B. C. (2009), 'Mental accounting in portfolio choice: Evidence from a flypaper effect', *American Economic Review* **99**(5), 2085–2095.

Deary, I., Strand, S., Smith, P. & Fernandes, C. (2007), 'Intelligence and educational achievement', *Intelligence* **35**, 13–21.

Edmonds, E. (2002), 'Reconsidering the labeling effect for child benefits: Evidence from a transition economy', *Economics Letters* **76**(3), 303–309.

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

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.

Frey, M. C. & Detterman, D. K. (2004), 'Scholastic Assessment or *g*?', *Psychological Science* **15**(6), 373–378.

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 evaluation periods', *Quarterly Journal of Economics* **112**(2), 631–645.

Greiner, B. (2004), 'An online recruitment system for economic experiments', *Forschung und wissenschaftliches Rechnen* **63**, 79–93.

Groenland, E. & Schoormans, J. (1994), 'Comparing mood-induction and affective conditioning as mechanisms influencing product evaluation and product choice', *Psychology and Marketing* **11**(2), 183–197.

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

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

Kahn, B. E. & Isen, A. M. (1993), 'The influence of positive affect on variety seeking among safe, enjoyable products', *Journal of Consumer Research* **20**(2), 257–270.

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.

Lewinsohn, S. & Mano, H. (1993), 'Multi-attribute choice and affect: The influence of naturally occurring and manipulated moods on choice processes', *Journal of Behavioral Decision Making* **6**(1), 33–51.

Milkman, K. L. & Beshears, J. (2009), 'Mental accounting and small windfall gains: Evidence from an online grocer', *Journal of Economic Behavior and Organization* **71**(2), 384–394.

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.

Plott, C. R. (1997), 'Laboratory experimental testbeds: Application to the PCS auction', *Journal of Economics & Management Strategy* **6**(3), 605–638.

Qiu, C. & Yeung, C. (2007), 'Mood and comparative judgment: Does mood influence everything and finally nothing?', *Journal of Consumer Research* **34**(5), 657–669.

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

Rabin, M. & Weizsäcker, G. (2009), 'Narrow bracketing and dominated choices', *American Economic Review* **99**(4), 1508–1543.

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

Roth, A. E. (2002), 'The economist as engineer: Game theory, experimentation, and computation as tools for design economics', *Econometrica* **70**(4), 1341–1378.

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

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

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

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

Thaler, R. H. & Sunstein, C. R. (2003), 'Libertarian paternalism', *American Economic Review* **93**(2), 175–179.

Thaler, R. H., 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.

Winkielman, P., Berridge, K. & Wilbarger, J. (2005), 'Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value', *Personality and Social Psychology Bulletin* **31**(1), 121.

A Additional regression tables

Table A.1: Treatment Effects on Volume of Consumed Beverages

Dependent variable: Consumed beverage volume (in liter per person)

	Full sample			Restricted sample		
	(1)	(2)	(3)	(4)	(5)	(6)
1 if Either voucher	-0.021	-0.068*	-0.077*	0.022	-0.005	-0.015
	(0.041)	(0.035)	(0.043)	(0.046)	(0.044)	(0.046)
1 if Label treatment	0.050	0.074*	0.078*	-0.014	0.006	0.001
	(0.049)	(0.042)	(0.043)	(0.052)	(0.051)	(0.050)
Meal consumption		0.013***	0.013***		0.008**	0.008**
		(0.003)	(0.003)		(0.003)	(0.003)
Outside temperature (in °C)			-0.000			-0.001
			(0.004)			(0.004)
Controls for day of the week	No	No	Yes	No	No	Yes
Constant	0.828***	0.506***	0.511***	0.889***	0.686***	0.736***
	(0.017)	(0.064)	(0.101)	(0.017)	(0.084)	(0.105)
N.Obs.	187	187	187	157	157	157

Notes: OLS estimates. Columns 1-3 report results for the full sample; columns 4-6 report results for a sample from which potentially distorted participants are excluded (see text for details). Robust standard errors clustered on day are in parentheses. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Table A.2: Treatment Effects on Average Price of Consumed Beverages

Dependent variable: Average price of consumed beverages (in euro per liter)

	Full sample			Restricted sample		
	(1)	(2)	(3)	(4)	(5)	(6)
1 if Either voucher	-0.623	-1.488	-0.633	-0.482	-1.114	-0.638
	(1.106)	(1.065)	(1.361)	(1.067)	(1.088)	(1.517)
1 if Label treatment	3.520**	3.971**	3.895***	3.883***	4.353***	4.107***
	(1.555)	(1.474)	(1.177)	(1.141)	(1.180)	(1.112)
Meal consumption		0.246***	0.243***		0.185**	0.186***
		(0.048)	(0.048)		(0.067)	(0.066)
Outside temperature (in °C)			-0.263*			-0.178
			(0.148)			(0.172)
Controls for day of the week	No	No	Yes	No	No	Yes
Constant	19.014***	13.065***	17.143***	19.978***	15.221***	18.740***
	(0.660)	(1.134)	(2.467)	(0.787)	(1.759)	(2.972)
N.Obs.	186	186	186	157	157	157

Notes: OLS estimates. Columns 1-3 report results for the full sample; columns 4-6 report results for a sample from which potentially distorted participants are excluded (see text for details). One table in the Cash Treatment did not consume any beverages and is excluded from this analysis. Robust standard errors clustered on day are in parentheses. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

B Instructions

The general instructions (shown below) were identical for both treatments. The only differences between treatments appeared in the on-screen instructions that were shown before the main stage of the experiment. The on-screen instructions are documented after the general instructions.

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 euros

In addition, you receive **2.50 euros** for having showed up on time. The 2.50 euros 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:

$$\text{Total price of purchase} = (\text{units of housing} \times \text{price per unit of housing}) + (\text{units of clothing} \times \text{price per unit of clothing})$$

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

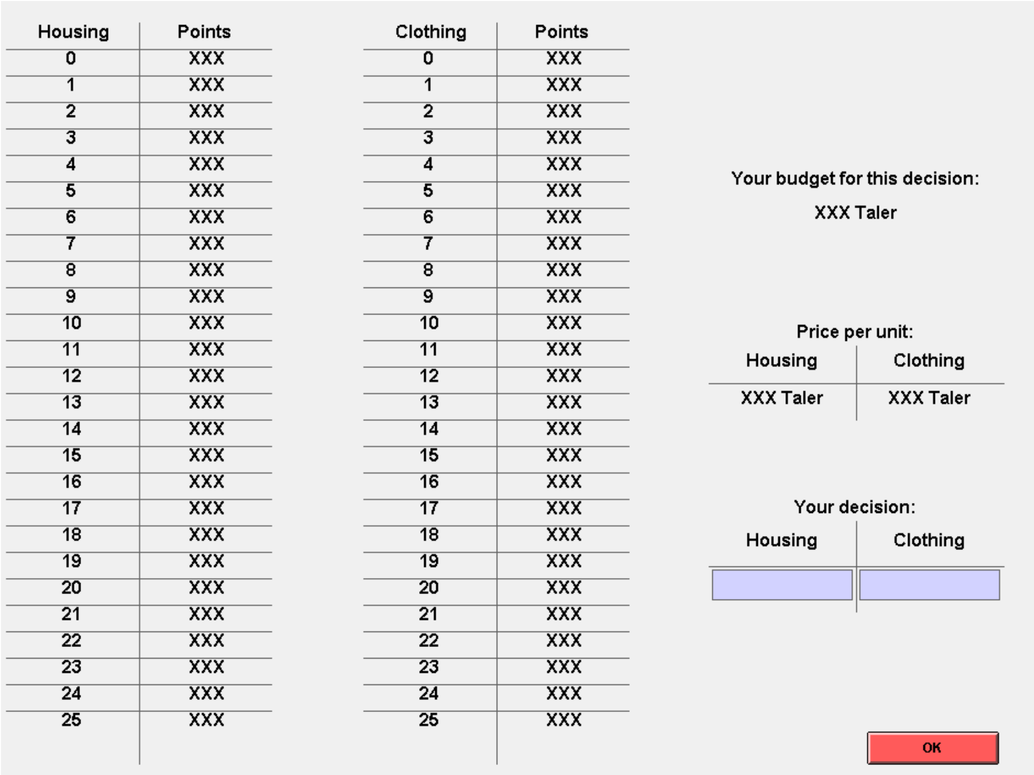
Example of a purchase

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

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**.